EXPANDING THE GENETIC CODE TO PROBE BIOLOGICAL SYSTEMS

A Dissertation

by

ALFRED ANTHONY TULEY

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Wenshe R. Liu
Committee Members,	Tadhg Begley
	Frank Raushel
	Weston Porter
Head of Department,	Francois Gabbai

December 2015

Major Subject: Chemistry

ABSTRACT

The chemical modification of proteins has been a longstanding interest in the scientific community. In addition to the natural modifications necessary for life to function, unnatural covalent modifications are particularly useful because they facilitate research efforts that require the precise manipulation of protein, including the installation of fluorescent labels, post-translational modification mimics, and affinity reagents. Historically, appending such modifications onto proteins was achieved by generating covalent adducts onto one of the twenty canonical amino acids. However, such modifications are not site-selective, and may interfere with the native function of the modified protein.

Genetic code expansion can overcome the limitations inherent to canonical amino acid modification, especially when bioorthogonal functional groups are incorporated. Using orthogonal aminoacyl-tRNA synthetase-tRNA pairs, one can reliably obtain homogenous samples of modified protein in a site-selective manner. In order to fully understand the steric requirements of a rationally designed pyrrolysyl-tRNA synthetase, several large *meta*-substituted phenylalanine derivatives were synthesized and incorporated into superfolder green fluorescent protein using this synthetase. All synthesized substrates were incorporated, albeit with differing incorporation efficiencies. Moreover, this synthetase was found to incorporate 3-formyl-phenylalanine, an aldehyde-based amino acid that can be directly installed onto proteins. Prior to this work,

only indirect post-translational approaches could install aldehydes, and these methodologies were limited as to where the modification could occur. Aldehyde labeling occurs rapidly at neutral pH, and peptide cyclization has been accomplished using the aldehyde to form a thiazolidine linkage with an N-terminal cysteine, further demonstrating the rich chemistry available to aldehydes. Finally, efforts to optimize azidophenylalanine bioconjugation led to a kinetic investigation of copper-catalyzed click chemistry, which led to a revision of the currently accepted mechanism; it is proposed that copper-catalyzed click chemistry requires two separate copper-chelating events, with one equivalent of copper binding to azide, and another equivalent of copper binding to alkyne. The work presented herein demonstrates that phenylalanine derivatives are useful substrates for probing and manipulating biological systems, as well as providing opportunities for discovery in chemical biology.

DEDICATION

This work is dedicated to the loving grandparents I have lost during my graduate studies: Alfred Dewey Tuley, Ladelle Tuley, and Elia Palos (affectionately known as Mamu). I miss them all terribly, and I love them very much.

ACKNOWLEDGEMENTS

Going to graduate school and obtaining my PhD has been a challenging ordeal for me, and I would not have been able to survive without the love and support of my family and friends. My parents have been a constant source of inspiration, and instilled a love of science and learning at a very early age. There is no way that I could ever repay them for the endless love and support they have given me throughout my life. I am sincerely grateful to have the wonderful Jennifer Stark in my life. She is the love of my life, and I thank her for being there for me during the toughest times and for going on so many adventures with me. Sharing this life with her has been a true blessing. My entire extended family has been a wellspring of support. Thank you John and Susan, Katie, Grandma Pat, John D., Grandpa Al, Grandma Ladelle, Mamu and Popo, Julian, Sarah, Tony and Tina, Ronnie and Suzie, Ron, Oscar, and Victor. Thank you to all my friends near and abroad that have helped me unwind and have fun when I really needed it. Very special thanks to my good friend Mikail Abbasov, who has been a constant source of inspiration and someone who I could rely on for much-needed sanity breaks. My heartfelt thanks to Jennifer Stark, Jeffery Tharp, and Vangmayee Sharma for proofreading this dissertation and providing much-needed feedback.

I would like to extend my sincerest gratitude to my research advisor, Prof. Wenshe Liu, for all of his support and encouragement throughout the years. His creativity and vision have been inspirational, and I am eternally grateful for the opportunity to work in his laboratory. I would also like to thank my committee members

for their support over these past few years. Thanks to Prof. Tadhg Begley for his unyielding enthusiasm for science and for his helpful advice throughout my graduate career. Thanks to Prof. Frank Raushel for the extraordinary enzymology lectures that I was fortunate enough to attend. Thank you Prof. Weston Porter for challenging me to consider the "big picture" and to look at my research through a different perspective. Thank you to all of the professors who have played a role in my graduate education: Profs. Daniel Singleton, Jiong Yang, John Gladysz, Daniel Romo, Brian Connel, and Kenn Harding.

I am also thankful for all of my fellow group members, past and present, that have helped me become the scientist I am today. Very special thanks to Dr. Yane-Shih "Eric" Wang, Dr. Catrina Reed, Dr. Yadagiri Kurra, Dr. Bo Wu, Dr. Yu Zeng, Keturah Odoi, Yan-Jiun Lee, and Jeffery Tharp for extensive guidance and numerous scientific discussions throughout my graduate career. Thanks to Dr. Morgan Shirley, Dr. Omar Robles, Dr. Gang Liu, Dr. Carolyn Leverett, and Dr. JC Reyes for their continued support during my time in the Romo group and for strengthening my knowledge in synthetic organic chemistry. Lastly, I would be remiss if I neglected to thank Prof. Daniel Romo for his support during my first two years of graduate school, and for his generous support and encouragement when I decided to alter my career path.

NOMENCLATURE

AIBN	Azobisisobutyrylnitrile
Amp	Ampicillin
AzF	Azidophenylalanine
AzFRS	Azidophenylalanyl-tRNA synthetase
BTTAA	bis[(*tert*-butyltriazoyl)methyl]-[(2-carboxymethyltriazoyl)methyl]-amine
BTTP	3-[4-({bis[(1-tert-butyl-1H-1,2,3-triazol-4-yl)methyl]amino}methyl)-1H-1,2,3-triazol-1-yl]propanol
CAT	Chloramphenicol acyltransferase
cfu	Colony forming unit
Cm	Chloramphenicol
CuAAC	Copper-catalyzed azide-alkyne cycloaddition
CV	Column volume
DBCO	Dibenzocyclooctyne
DCM	Dichloromethane
DEAD	Diethyl azodicarboxylate
DIBAL-H	Diisobutylaluminum hydride
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide

DNA	Deoxyribonucleic acid
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EPL	Expressed protein ligation
ESI-MS	Electrospray ionization-mass spectrometry
EtOAc	Ethyl acetate
FRET	Förster resonance energy transfer
Hex	Hexanes
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kan	Kanamycin
MeCN	Acetonitrile
MjTyrRS	*Methanococcus jannaschii* tyrosyl-tRNA synthetase
mRNA	Messenger ribonucleic acid
NBS	N-bromosuccinimide
NCAA	Non-canonical amino acid
NCL	Native chemical ligation
NMR	Nuclear magnetic resonance
O.D.	Optical density
OmpX	Outer membrane protein X
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PheRS	Phenylalanyl-tRNA synthetase

ProY	Propargyl-tyrosine
PTM	Post-translational modification
Pyl	Pyrrolysine
PylRS	Pyrrolysyl-tRNA synthetase
r.p.m.	Rotations per minute
RNA	Ribonucleic acid
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
sfGFP	Superfolder green fluorescent protein
SPPS	Solid phase peptide synthesis
TEV	Tobacco etch virus
THF	Tetrahydrofuran
tRNA	Transfer ribonucleic acid
$tRNA^{Pyl/CUA}$	Pyrrolysine tRNA (anticodon: CUA)
$tRNA^{Tyr/CUA}$	Tyrosine tRNA (anticodon: CUA)

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

1.1. Protein Biosynthesis and Post-Translational Modifications

The twenty canonical amino acids are a universal necessity for life to function [1]. Through the judicious and practically inerrant translation of the genetic code, organisms synthesize a wide variety of proteins and enzymes, biomolecules essential for catalyzing biochemical processes or for providing essential structural motifs in the organism. Each amino acid possesses a unique functional group substituted at the β-position relative to the carboxylic acid, and the summation of these functional groups in a polypeptide dictate the specific three-dimensional structure and properties of a protein.

The central dogma of biology outlines the general method of information transfer (Figure **1.1**);[2] DNA is transcribed into messenger RNA (mRNA), which is finally translated into proteins. To put succinctly, during the translation process, elongation factors deliver aminoacylated transfer RNA (tRNA) to the A site of the ribosome, a biomolecule responsible for coupling amino acids to the growing polypeptide chain. Upon addition of this amino acid via peptidyltransferase activity, the ribosome is moved toward the 3' end of mRNA, moving one codon length (three nucleotides) at a time in the process, ejecting uncharged tRNA from the E site of the ribosome and making the A site vacant for the next aminoacylated tRNA. Translation continues until the ribosome encounters a stop codon, one of three unique nucleotide sequences (UAG, UAA, and UGA) that recruit release factors, polypeptides that mimic the three-dimensional structure of tRNA and trigger disassembly of the translational machinery.

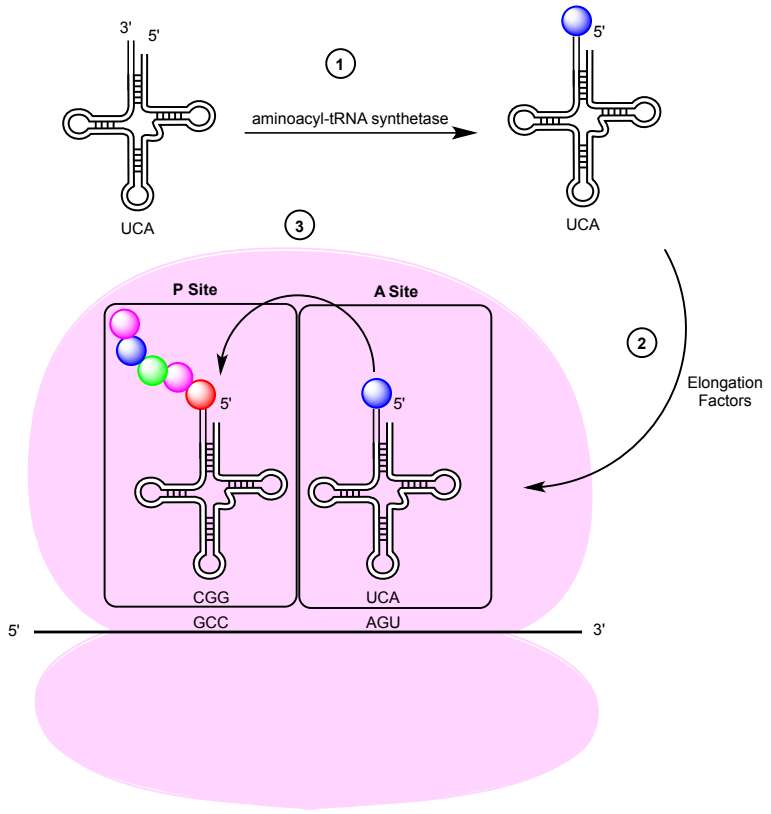


Figure 1.1. Protein translation. (1) tRNA is aminoacylated by aminoacyl-tRNA synthetase, followed by (2) elongation factor-mediated transport into the ribosome (in pink) and (3) elongation of the nascent peptide chain via peptidyltransferase activity of the ribosome.

After translation, proteins may require further processing to become fully functional; amino acid residues are often chemically modified to alter their properties.[3] These post-translational modifications (PTMs) are capable of adding or attenuating charges (i.e. acetylation, methylation, phosphorylation),[3,4] modulating transcription factor activities (glycosylation),[5-7] and/or appending signals for protein trafficking and degradation (e.g. ubiquitination).[8,9] The summation of various PTMs alter the form and

function of proteins by modulating their affinity for other biomolecules as well as providing localized perturbations that can alter the secondary structure of said protein. A survey of common post-translational modifications is provided in Figure **1.2**.

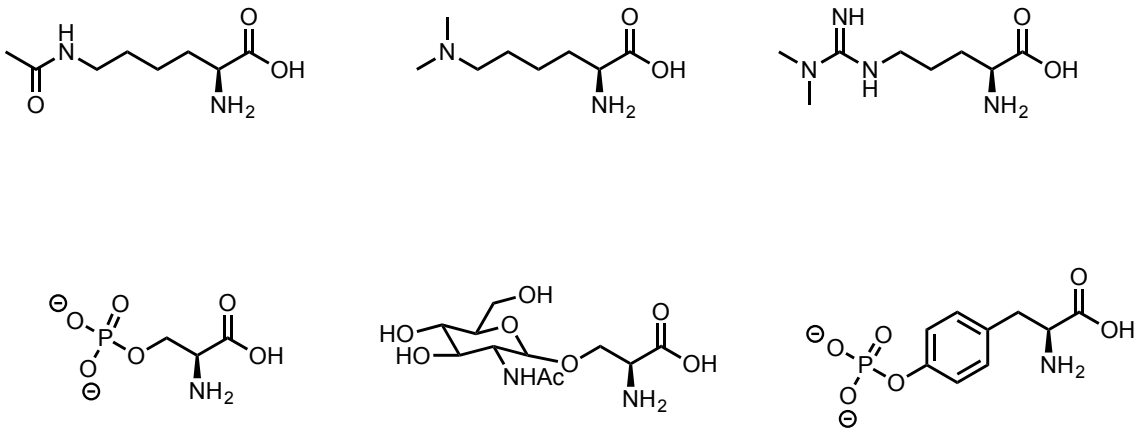
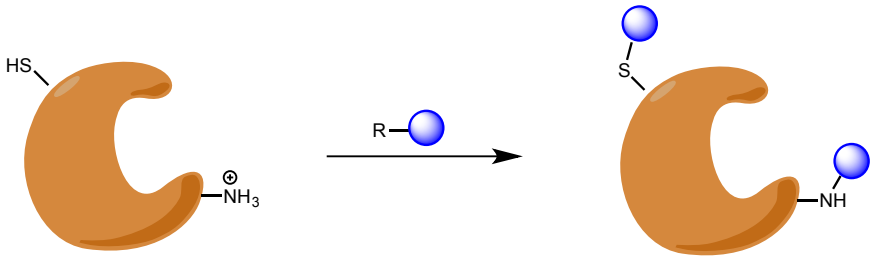


Figure 1.2. Selected examples of post-translational modifications found in nature.

The existence of post-translational modifications verify that "default" amino acid residues are not all that is required for life to function, and that small changes can have a profound influence on the overall form and function of a protein. Unnatural chemical modification of proteins can serve a variety of purposes, including the installation of PTM mimics, spectroscopic probes, and affinity tags. Initially, investigators that sought to chemically modify proteins were limited to modification onto one of the twenty canonical amino acids.[10-14] Of these, the two most commonly employed substrates were cysteine and lysine. Cysteine has been a useful handle for protein modification because

it is relatively rare and is highly nucleophilic, thus providing opportunities for selective modifications. Proteins have been successfully modified at cysteine residues using a number of electrophiles, including disulfides, iodoacetamides, and Michael acceptors. Lysine has also been utilized for bioconjugation purposes, and although this amino acid is far more abundant compared to cysteine, the number of methods available for chemoselective primary amine modification has led to the modification of lysine in numerous methodologies. Activated esters, sulfonyl chlorides, isocyanates, and isothiocyanates have all been useful electrophiles for lysine modification.



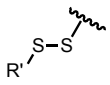
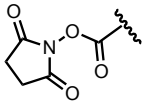
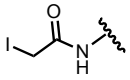
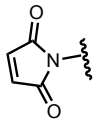
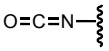
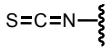
Residue	Reagent Electrophile (R)	Residue	Reagent Electrophile (R)
Cys		Lys	
Cys		Lys	
Cys		Lys	
		Lys	

Figure 1.3. Bioconjugation methods using cysteine and lysine. Figure adapted from Sletten and Bertozzi.

While the above methodology has utility under the appropriate context, modification of native amino acids has several drawbacks. For example, modification is usually not site-selective, unless one can ensure that only one residue of the amino acid is present in the protein. Secondly, the modified residues may be essential for the protein to function properly. Lastly, their utility as handles for *in vivo* experiments is diminished because of the organism-wide presence of cysteine, lysine, and biomolecules with similar functionality.

1.2. Bioorthogonal Chemistry

In order to circumvent the limitations imposed upon canonical amino acid bioconjugation, it is necessary to utilize chemoselective reactions that do not interfere with functional groups found in organisms. Such reactions are said to be "bioorthogonal",[14] and several notable examples have been developed and utilized for various applications. All reported bioorthogonal reactions have been simple, second-order reactions that can react relatively quickly and selectively under biological settings. As discussed later in this Chapter, bioorthogonal reagents can be installed using a number of different methods, including incorporation as non-canonical amino acids. The most common bioorthogonal reactions are discussed in detail below (Figure **1.4**).

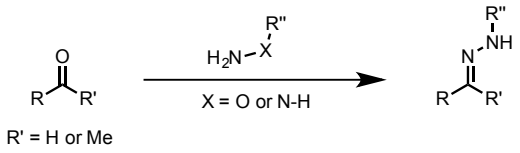
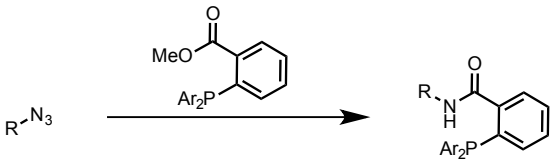
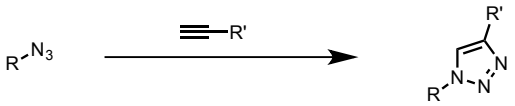
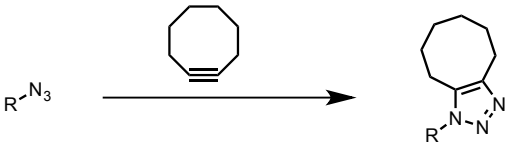
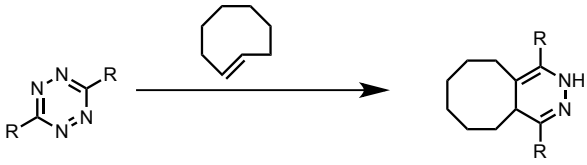
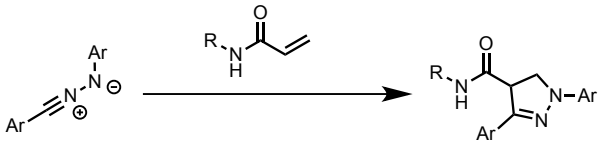
Entry	General Reaction Scheme	Comments
1	 R' = H or Me; X = O or N-H	Acidic conditions often required. Susceptible to hydrolysis unless Pictet-Spengler adduct
2		Traceless variants have been developed. Phosphine reagents are prone to oxidation
3		Copper (I) required, but is toxic to organisms
4		No copper required. Cyclooctynes can also undergo cycloaddition with nitrones
5		Extremely rapid reaction. Tetrazines also react well with norborenes and cyclopropenes
6		Nitrile imine generated *in situ*. Unstrained olefins are capable of reacting

Figure 1.4. Survey of common bioorthogonal reactions.

1.2.1. Carbonyl-Based Methods

The earliest known example of a bioorthogonal reaction is the utilization of aldehydes and ketones for condensation with hydrazines or hydroxylamines, and this reaction continues to be a useful bioconjugation tool[15-17] (Figure **1.4**, entry 1). Chemical oxidation can occur on endogenous 1,2-diols or similar functional groups,[18-21] which can then be modified using one of the α-effect nucleophiles. Although carbonyl-based bioconjugation reagents are easily synthesized, there are some disadvantages to this technique. The bioorthogonality of the reaction is highly dependent on context; although aldehydes and ketones do not exist on the surface of proteins, they are found in a number of other natural biomolecules that can lower the efficiency and selectivity of the desired reaction.[14] Secondly, bioconjugation with ketones require acidic conditions for efficient labeling, which may not be amenable to more sensitive systems. Lastly, the resulting oximes and hydrazones have finite stability, as hydrolysis will occur over time.[22] However, it should be noted that Bertozzi was able to overcome hydrolytic instability by trapping the resulting oxime as a Pictet-Spengler product.[23]

1.2.2. Staudinger Ligation

Ideally, the functional groups necessary for a bioorthogonal reaction are completely absent from biological systems. For this reason, reactions involving azides have seen widespread applications for bioconjugation methodologies. In addition to their absence in native biomolecules,[24] azides are small and relatively inert, meaning that their installation would result in minimal perturbation of a system.[25,26] The first azide-based strategy to be reported is the Staudinger ligation, a variant of the classic Staudinger

reaction.[27] Typically, azides will undergo a Staudinger reaction with phosphine reagents to yield amines. However, using specialized phosphine reagents, azides can be converted into amides via trapping of the aza-ylide with a nearby ester group. Traceless forms of this reaction have also been developed.[28-31] It should be noted that azides are capable of being reduced by cellular thiols, but the reaction rate is slow. One major disadvantage of this technique is that phosphine reagents are susceptible to oxidation.

1.2.3. Azide-Alkyne Cycloaddition

Perhaps the most commonly employed bioorthogonal reaction is the copper-catalyzed azide-alkyne cycloaddition, commonly abbreviated as CuAAC. In the presence of copper (I) species, azides and alkynes undergo a 1,3-dipolar cycloaddition to generate a stable triazole product.[32,33] The reaction rate for this reaction is relatively fast, and occurs under mild conditions. Due to the robust nature of this reaction, it has been utilized in a number of other applications throughout several disciplines of chemistry.[34-38]

Unfortunately, copper is toxic to biological systems, but recent reports have utilized copper-chelating ligands to reduce copper toxicity.[39,40] Alternatively, a copper-free azide-alkyne cycloaddition has also been developed.[41] Rather than lowering the transition state energy of the reaction, copper-free cycloaddition occurs by raising the energy of the starting material, a feat achieved using highly strained cyclooctyne reagents. The reaction is quite useful for *in vivo* applications, but the reaction is much slower than CuAAC and the cyclooctyne reagents are more difficult to synthesize, compared to the simple terminal alkynes and primary azides necessary for CuAAC.

1.2.4. Tetrazine Ligation

One of the fastest available bioorthogonal strategies is the inverse electron demand cycloaddition with tetrazines and strained alkenes, such as *trans*-cyclooctenes, cyclopropenes, and norborenes.[42-46] The initial cycloaddition is followed by expulsion of nitrogen, yielding a stable cyclic product. The rapid nature of this reaction with cyclooctenes has resulted in its widespread utilization for *in vivo* applications. However, although the rapid nature of the reaction is useful, the reagents are difficult to synthesize and are consequently inaccessible to laboratories that lack experience in synthetic organic chemistry.

1.3. Methods to Install Non-Canonical Amino Acids

As mentioned above, the chemical modification of proteins is a naturally occurring phenomenon that results in altered reactivity and specificity. The study of post-translational modifications has been historically difficult due to the dynamic nature of protein modifications, and the fact that these changes are relatively small, leading to difficult or impossible purification of proteins with the desired modifications. Additionally, chemical modification to bestow non-native functionality onto proteins can be useful, but the approach has been hindered by a dearth of methods that are chemoselective and site-selective. However, the advent of bioorthogonal chemistry has given chemical biologists the necessary tools for achieving selectivity in the presence of numerous competing biomolecules.

1.3.1. Peptide ligation

Prior to the development of genetic code expansion, researchers were limited to protein samples obtained via indirect, *in vitro* methodologies. The advent of solid phase peptide synthesis (SPPS)[47] allowed for the convenient synthesis of peptides without regards to limitations imposed by the protein translational machinery. Consequently, oligopeptides could be reliably synthesized that contained PTMs or other "unnatural" functional groups. In conjunction with native chemical ligation (NCL)[48], these oligopeptides could be conjugated to the N-terminus of another peptide fragment to eventually acquire full-length protein containing a desired modification. In tandem with expressed protein ligation (EPL)[49], homogenous samples of modified protein became far more accessible (Figure **1.5**). Despite the accomplishments made through SPPS, NCL, and EPL, the techniques are labor intensive and are limited with regard to where the protein modification can be installed; this is largely a limitation from SPPS, which can only synthesize peptides of a certain length before becoming prohibitively inefficient.

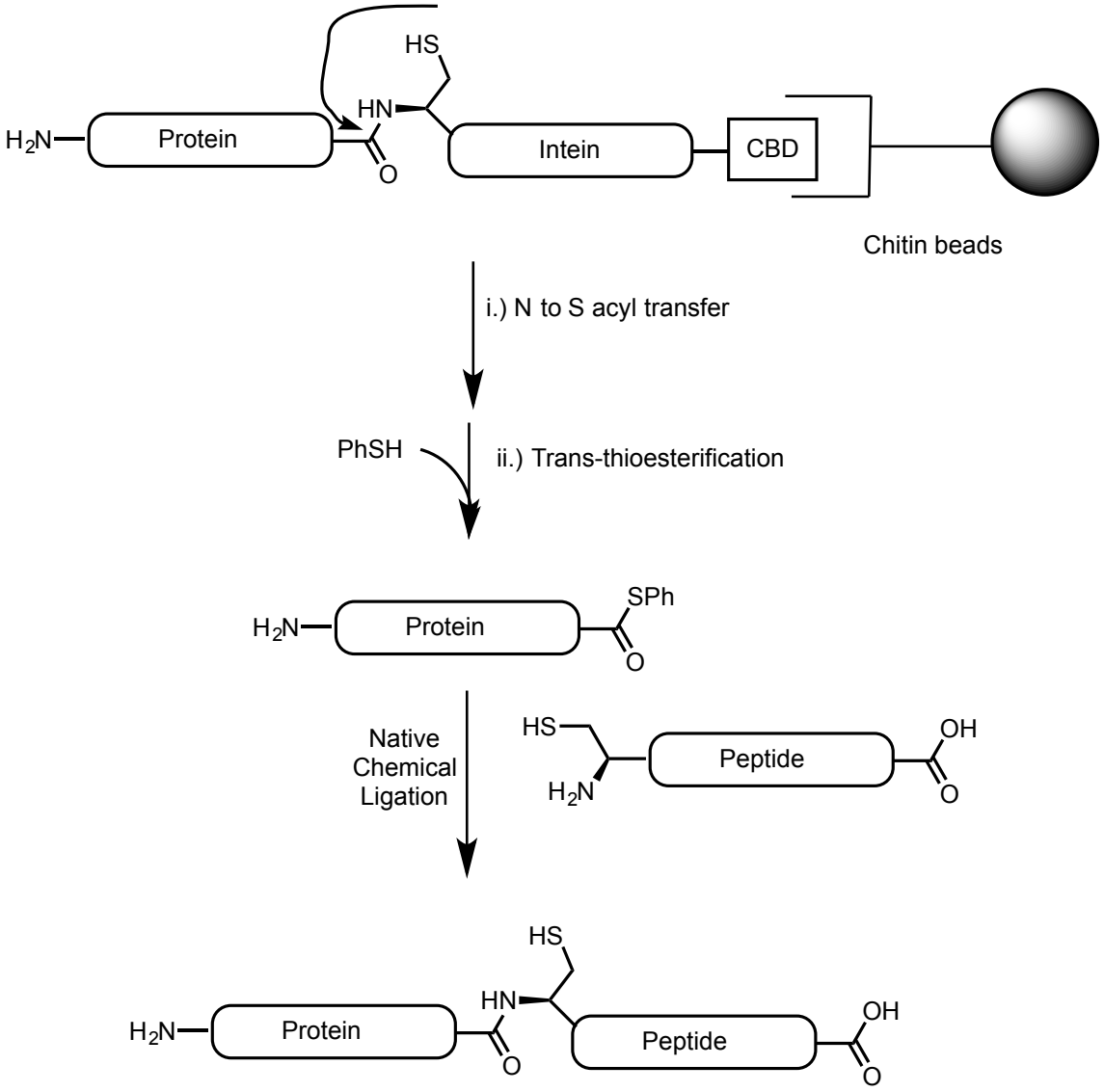


Figure 1.5. Expressed protein ligation (EPL). Native chemical ligation (NCL) is the last step.

1.3.2. Genetic Code Expansion

A great deal of effort has been put forth in order to exploit the native translational machinery for NCAA incorporation. An important observation about protein translation is that codons and amino acids are not directly linked, but that the anticodon region of the tRNA is crucial for interaction with mRNA codons (refer to Figure **1.1**). In other words, a tRNA which is misaminoacylated will dutifully install the amino acid in question solely based on whenever its appropriate codon appears in the mRNA sequence, regardless of which particular amino acid is installed.[50,51] Early efforts to incorporate unnatural amino acids utilized chemically acylated tRNAs which could be used *in vitro*, but the approach suffered from technical difficulties such as low yields, hydrolysis, and competition with endogenous amino acids which require the same codon.[52-57] Alternatively, *in vivo* incorporation could be accomplished using canonical machinery, but such an approach requires amino acids that must be structurally similar to canonical amino acids, and competition between endogenous substrates still occurs unless auxotrophic strains are utilized with strictly defined media.[58-61] Furthermore, the unnatural amino acid will be incorporated at all sites in which the canonical amino acid would be coded, thereby limiting the finesse of such an approach. Selected examples of amino acid analogues that have been metabolically incorporated are shown in Figure **1.6**; of the analogues successfully incorporated, methionine surrogates are particularly useful for the global installation of bioorthogonal groups.[62]

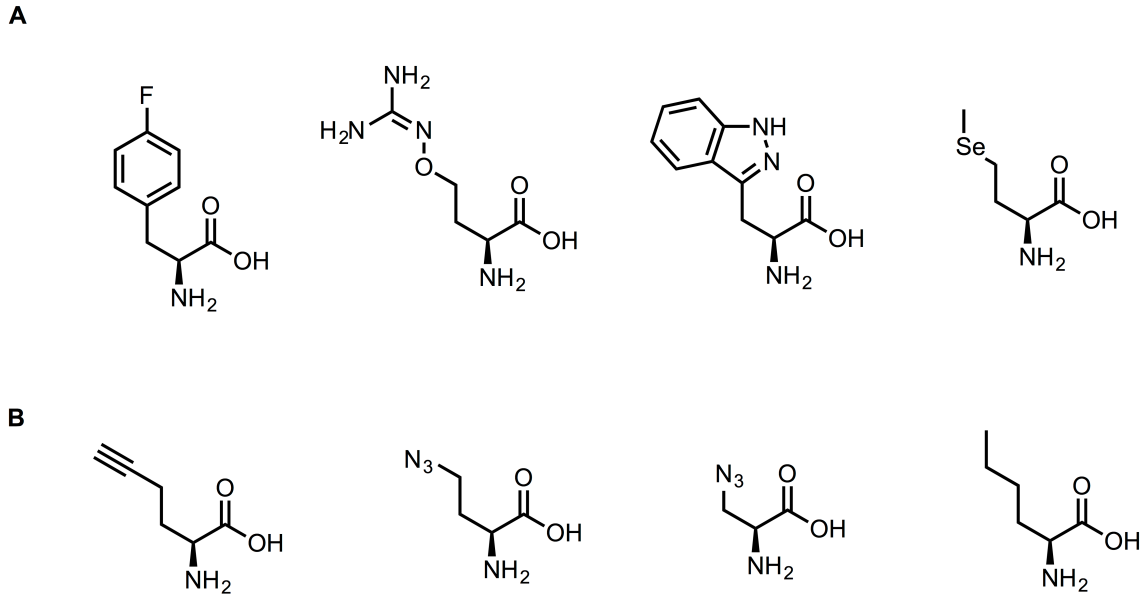


Figure 1.6. Non-canonical amino acids incorporated using native machinery. **A.** Early examples. **B.** Methionine analogues incorporated by Tirrell and Bertozzi.

The laboratory of Peter Schultz reported the first successful, site-specific *in vivo* incorporation of unnatural amino acids in 2001.[63] Using a tyrosyl tRNA synthetase-tRNA pair from *Methanococcus jannaschii*,[64] the Schultz laboratory was able to expand the genetic code in *Escherichia coli* with the non-canonical amino acid O-methyl-L-tyrosine at UAG codon sites. The MjTyrRS-tRNA$^{Tyr/CUA}$ pair was optimized for O-methyl-L-tyrosine using two directed evolution approaches, one to optimize the tRNA and another for optimization of the synthetase. To identify an efficient tRNA mutant for methyltyrosine incorporation, a randomized library was first subjected to a round of negative selection, whereby tRNA capable of being charged by endogenous synthetases would die from toxic barnase expression. The survivors of this negative selection would

either be non-functional or orthogonal to synthetases native to *E. coli.* Plasmid DNA from the surviving library members was used to transform *E. coli* harboring a positive selection plasmid. Survivors from this round of selection were both functional and orthogonal to *E. coli* synthetases. From here, the chosen tRNA mutant was utilized for selection of a synthetase mutant capable of O-methyl-L-tyrosine incorporation. Cells possessing the synthetase library were grown in the presence of non-canonical amino acid and chloramphenicol (Cm); successful read-through of an in-frame amber codon in chloramphenicol acyltransferase (CAT) lead to cell survival in these conditions. Survivors were consequently capable of incorporating either the NCAA or any natural amino acids with the synthetase library. Replica plating in the presence and absence of non-canonical amino acid allowed for isolation of colonies that were only capable of growth in the presence of NCAA, and these survivors were further processed and resubjected to the selection conditions until an optimized, efficient synthetase mutant was isolated.

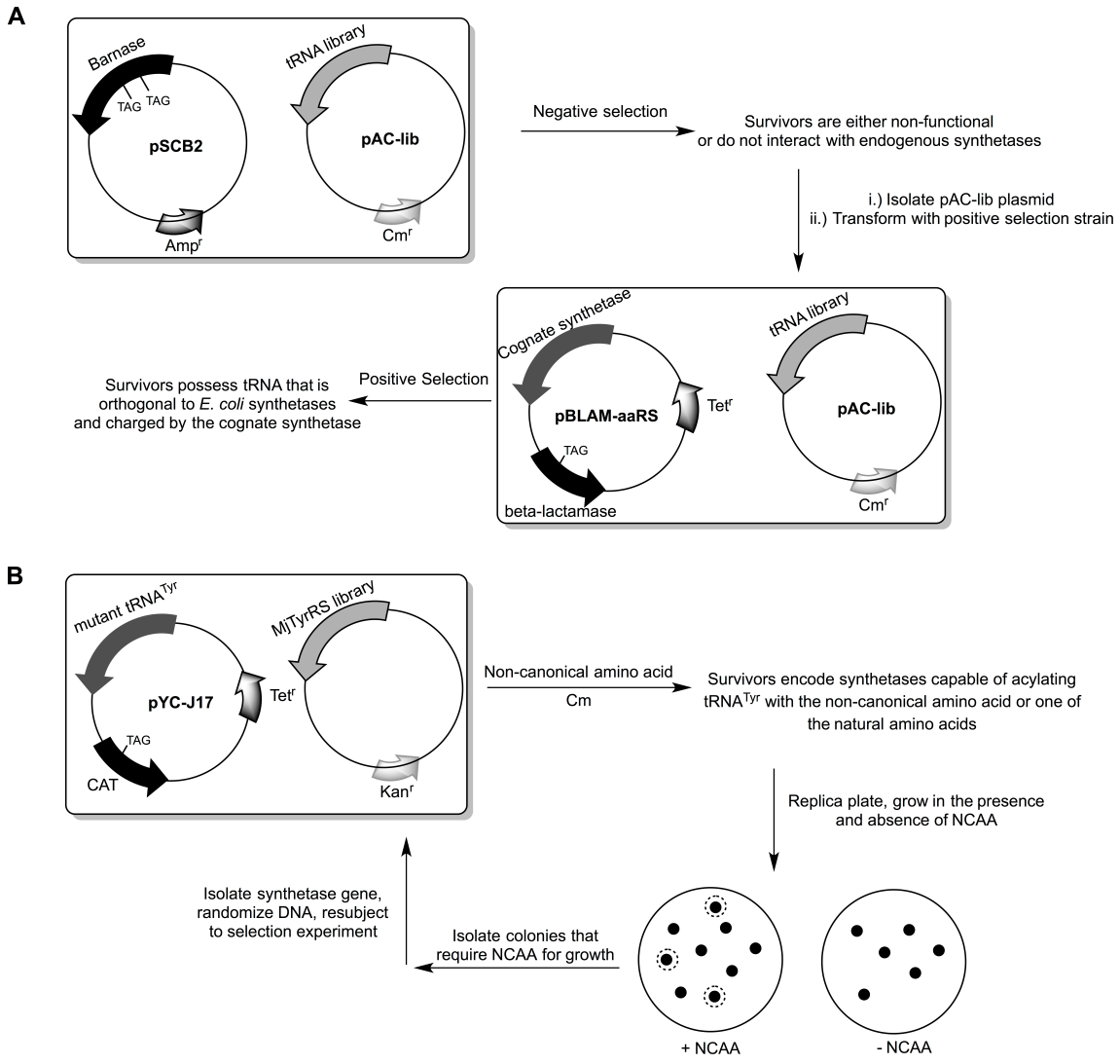


Figure 1.7. A. Schultz's method to evolve efficient, orthogonal $tRNA^{Tyr/CUA}$ mutants. **B.** Directed evolution to discover mutants capable of aminoacylating $tRNA^{Tyr/CUA}$ with O-methoxy-L-tyrosine.

In vivo utilization of the MjTyrRS-$tRNA^{Tyr/CUA}$ pair has been a successful strategy for a number of reasons. Considering that *M. jannaschii* is a methanogenic species of archaea, it belongs in a domain distinct from the prokaryotic *E. coli* and thus

has molecular machinery that is orthogonal in the latter system.[1] More specifically, the structural and recognition elements of *M. jannaschii* tRNA are distinct from those found in *E. coli*, which means the archaeal synthetase will not acylate endogenous tRNA. Further, the anticodon region of said tRNA is only loosely associated with MjTyrRS during translation, meaning that the anticodon region could be mutated without perturbing the affinity for the synthetase-tRNA pair. Also, MjTyrRS lacks an editing domain that may interfere with acylation of non-native substrates. Other requirements for successful genetic code expansion include successful cytoplasm transport and nonexistent acylation of the unnatural amino acid with native machinery. After evolving an amber-suppressing MjTyrRS-tRNA$^{Tyr/CUA}$ pair using randomized libraries, the discrepancies between native and orthogonal systems were enhanced, consequently boosting the fidelity of the system.

Following the initial report, a number of other non-canonical amino acids were incorporated using mutant forms of MjTyrRS. These amino acids included an IR-active probe for studying protein conformation,[65] reactive functionalities for bioconjugation and photocrosslinking,[66,67] and a photocaged tyrosine for light-responsive modulation of enzyme activity.[68]

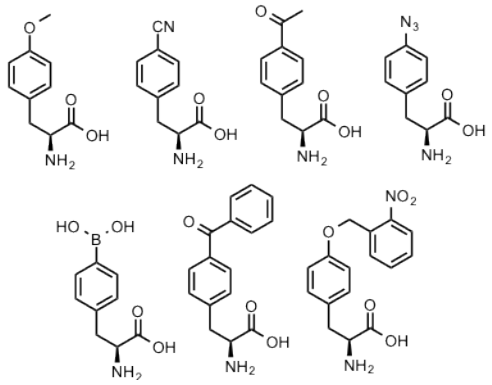


Figure 1.8. Selected examples of NCAAs incorporated with MjTyrRS mutants.

1.3.3. Exploiting Natural Instances of Genetic Code Expansion

In some instances, nature has devised methods to expand the genetic code via selective suppression of nonsense codons, namely the amber (UAG) and opal (UGA) stop codons. The 21st amino acid, selenocysteine, was discovered several decades ago and is incorporated via opal suppression; recognition of UGA as a "sense" codon instead of a stop codon requires a unique selenocysteine insertion sequence and elongation factor.[69] Because of these extraneous incorporation components, the utilization of selenocysteine machinery is inconvenient for genetic code expansion and its use has not been reported.

Yet another amino acid, pyrrolysine, was discovered in 2001 in methanogenic archaea, and was found to be incorporated via amber suppression.[70,71] With a unique pyrrolysyl-tRNA synthetase and pyrrolysyl tRNA, this amino acid is efficiently installed by reading through the UAG amber codon. The incorporation machinery is much

simpler than selenocysteine incorporation;[72] only the synthetase is needed, along with its cognate tRNA. Further, its delegation to only a few species of archaea and bacteria mean that the pyrrolysine machinery is orthogonal in many systems of interest, both prokaryotic and eukaryotic.

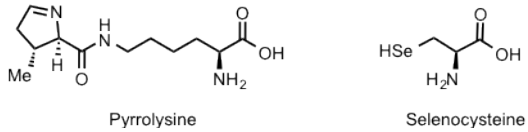


Figure 1.9. Structures of pyrrolysine and selenocysteine.

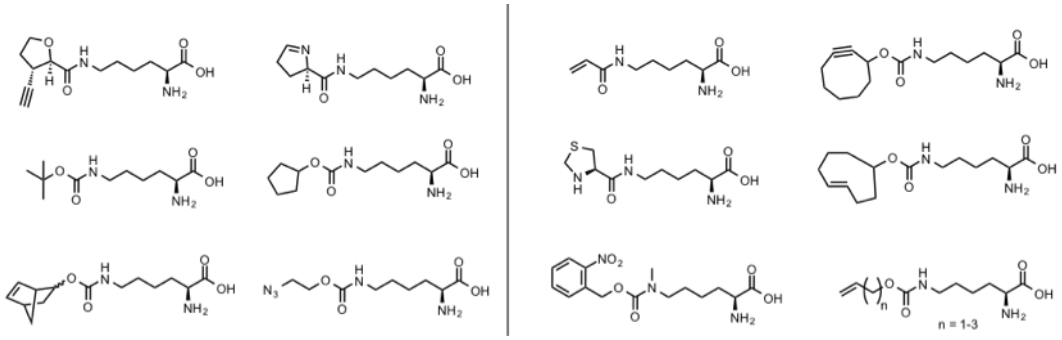


Figure 1.10. Amino acids incorporated using wild-type PylRS (left) and mutants of PylRS (right).

Due to the reasons stated above, the discovery of pyrrolysine and its incorporation machinery has been a boon to the field of non-canonical amino acids. Researchers first began to exploit the wild-type pyrrolysyl-tRNA-synthetase-

$tRNA^{Pyl/CUA}$ (PylRS-PylT) pair to incorporate analogues of pyrrolysine that possess functional groups amenable to bioorthogonal chemistry.[73-75] Shortly afterward, a number of research groups have disclosed the evolution of mutant PylRS-PylT pairs that are capable of incorporating a wider variety of substrates.[76,77] These non-canonical amino acids include Michael acceptors,[78] activated reagents for protein labeling,[79,80] and post-translational modification mimics.[81] Many more examples are present in the literature, and a comprehensive review has recently been published.[82]

1.3.4. Beyond Pyrrolysine Analogues: Altered Substrate Specificity of PylRS

Despite the advances made in genetic code expansion, the pyrrolysyl incorporation machinery had thus far been delegated to the incorporation of acylated lysine derivatives. Although these have been useful scaffolds for a number of applications, a wider variety of amino acid substrates would be desirable. In 2011, our group reported the directed evolution of a PylRS capable of incorporating phenylalanine and its derivatives, namely *p*-iodo and *p*-bromophenylalanine.[83] Prior to this discovery, the only technology capable of incorporating phenylalanine or tyrosine derivatives was the system devised by Schultz and coworkers. Considering that phenylalanyl-tRNA synthetase (PheRS) belongs to the same class as PylRS,[84] and that they share high structural homology, it seems reasonable that PylRS could be engineered to acylate phenylalanine and similar substrates.

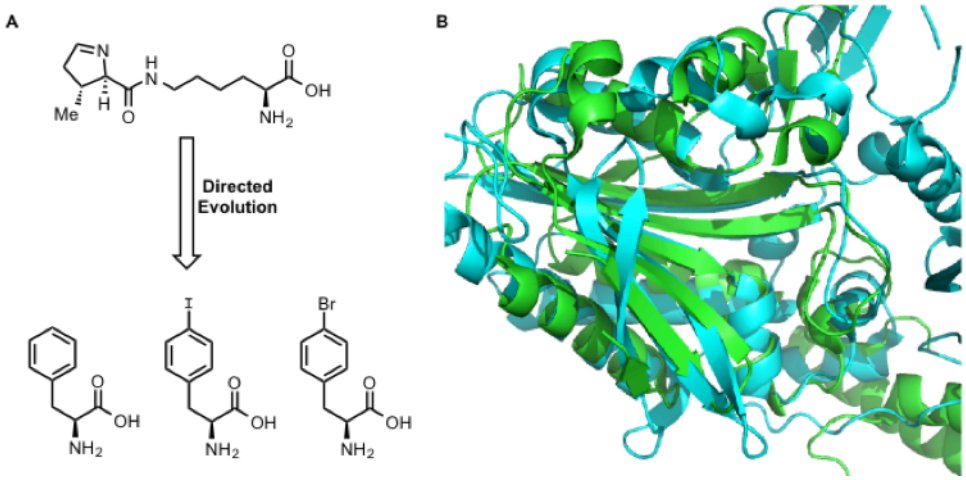


Figure 1.11.A. Directed evolution of PylRS alters the substrate specificity towards phenylalanine and its derivatives. **B.** Superimposed image of PylRS (2Q7G, green) and PheRS (3CMQ, teal)

1.4. Conclusions and Outlook

The discovery of the pyrrolysine incorporation machinery has proven itself to be an invaluable tool for discovery in chemical biology. Researchers are now in possession of an orthogonal system that naturally suppresses amber stop codons without extraneous recognition sequences, and its loose affinity for the native substrate can be easily exploited to suit the needs of a particular research project. The past decade has seen remarkable advancements in the field of genetic code expansion, and the scientific community now has a myriad number of options when research efforts require site-specific installation of spectroscopic probes, reactive functionalities, and post-translational modifications.

2. GENETIC INCORPORATION OF THIRTEEN NOVEL NON-CANONICAL AMINO ACIDS*

2.1. Introduction

Site-selective installation of non-canonical amino acids (NCAAs) at an amber codon is an efficient approach to synthesize proteins with unique functionalities; applications span from basic studies such as protein cellular localization and protein–protein interaction analysis, to biotechnological applications such as the synthesis of heat stable enzymes and therapeutic protein manufacturing.[85-89] Two aminoacyl-tRNA synthetase–tRNA pairs have been well adapted for the genetic incorporation of NCAAs at amber codons in bacteria. One is the tyrosyl-tRNA synthetase–$tRNA^{Tyr/CUA}$ pair that was derived from *Methanocaldococcus jannaschii*.[90-92] The other is the pyrrolysyl-tRNA synthetase (PylRS)–$tRNA^{Pyl/CUA}$ pair that naturally occurs in some methanogenic archaea.[93-96] Due to its broad-spectrum orthogonality from bacteria to human cells and the fact that it can be easily engineered to target a large variety of NCAAs, including natural amino acids with posttranslational modifications, the PylRS– $tRNA^{Pyl/CUA}$ pair has captivated researchers for the past several years.[97-111] One of our major contributions to the NCAA research field has been the development of PylRS mutants capable of incorporating a number of phenylalanine derivatives, which are substantially different from the structure of pyrrolysine, the native substrate of PylRS.[85,112,113] More specifically, we have recently shown that a rationally designed,

*Tuley, A.; Wang, Y.-S.; Fang, X.; Kurra, Y.; Rezenom, Y. H.; Liu, W. R. *Chemical Communications* **2014**, *50*, 2673. Reproduced by Permission of the Royal Society of Chemistry

N346A/C348A mutant of PylRS (PylRS(N346A/C348A)) is capable of incorporating seven *para*- and twelve *meta*-substituted phenylalanine derivatives at amber codons in coordination with $tRNA^{Pyl/CUA}$.[112,113] This broad substrate scope obviates the need to undergo the arduous task of discovering a new mutant for each NCAA. Herein we demonstrate that PylRS(N346A/C348A) has an even broader substrate scope than previously reported.

Our previous studies revealed a large active site pocket in PylRS(N346A/C348A).[112] Removal of the N346 side chain amide dismisses the steric clash that prevents the binding of the aromatic side chain of phenylalanine and the loss of the C348 thiol yields a cavernous pocket capable of binding the *para*- or *meta*-substituted phenylalanine described above. Interestingly, although phenylalanine derivatives with small *para*-substituents have shown to be ineffective substrates for PylRS(N346A/C348A), their isomers with *meta*-substituents act as highly efficient substrates of PylRS(N346A/ C348A) for their genetic incorporation at amber codons.[112,113] In other words, phenylalanine derivatives with *para*-substituents can only be incorporated when they possess large side chains. Upon further inspection, it appears that a majority of the vacancy in the active site pocket of PylRS(N346A/C348A) exists near the *meta* position of phenylalanine. Encouraged by our preliminary work, we reasoned that PylRS(N346A/C348A) could incorporate phenylalanine derivatives with more sterically demanding side chains.

2.2. Results and Discussion

Our investigation began with the synthesis and genetic incorporation of four different *para*-substituted phenylalanine derivatives (**1–4** in Figure **2.1.A**), each with a unique functionality and steric requirement. Synthesis of these derivatives followed the same strategy presented in one of our previous reports of the N346A/C348A mutant,[112] with the exception of NCAA **1**, which was synthesized using a different approach (refer to section 2.4.4 later in this Chapter). These four NCAAs were then tested for their tolerability by PylRS(N346A/C348A). An *E. coli* BL21(DE3) cell that harbors two plasmids, pEVOL-pylT-PylRSN346A/C348A and pET- pylT-sfGFP2TAG, was employed for the investigation. pEVOL-pylT- PylRSN346A/C348A contains genes coding PylRS(N346A/C348A) and $tRNA^{Pyl/CUA}$; pET-pylT-sfGFP2TAG carries a $tRNA^{Pyl/CUA}$ coding gene and a non-sequence-optimized superfolder green fluorescent protein (sfGFP) gene with an amber mutation in position S2 (sfGFP2TAG). The same cells were used in the initial test of the recognition of para-substituted phenylalanine derivatives by PylRS(N346A/C348A).[112] Growth in minimal media supplemented with 1 mM IPTG and 0.2% arabinose without NCAA afforded a minimal expression level of full-length sfGFP (<0.3 mg L^{-1}). Addition of any of **1–4** at 2 mM to the medium all promoted full-length sfGFP expression (Fig. 1A). The expression levels for **1–3** are comparable to that for *para*-propargyloxyphenylalanine (7.8 mg L^{-1}),[112] and the electrospray ionization mass spectrometry analysis of four purified sfGFP variants displayed molecular weights that agreed well with the theoretical values corresponding to full-length proteins with the first methionine (Figure **2.1.B**).

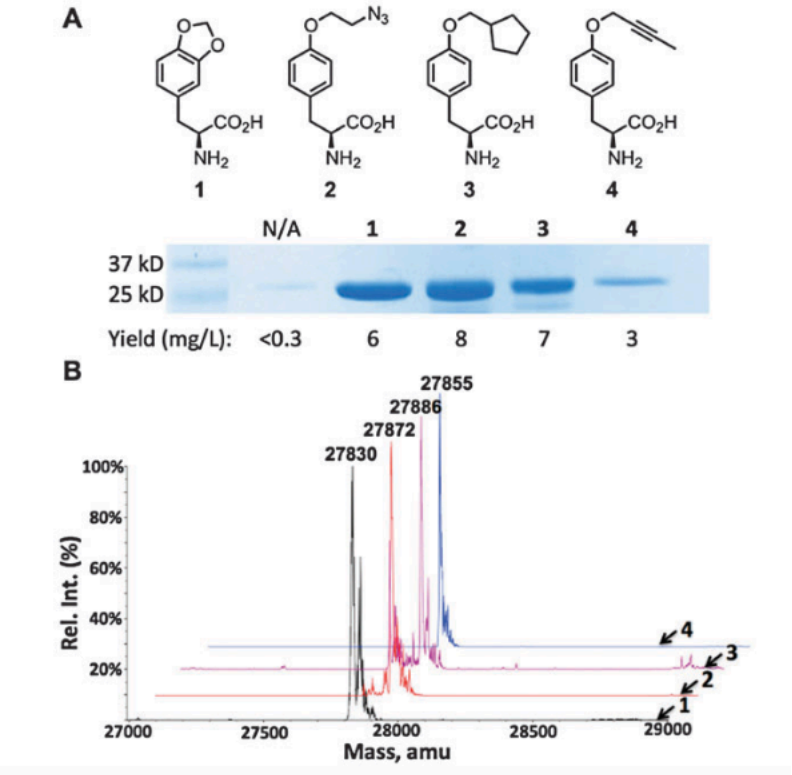


Figure 2.1. (A) Structures of **1–4** and their site-specific incorporation into sfGFP at its S2 position. (B) Deconvoluted ESI-MS spectra of sfGFP variants incorporated with **1–4**. Their theoretical values are 27 832 Da for **1**, 27 873 Da for **2**, 27 886 Da for **3**, and 27 856 Da for **4**. Satellite signals are largely due to metal ion adducts (i.e. Li, Na, K).

The results obtained for **1–4** demonstrate that PylRS(N346A/ C348A) tolerates phenylalanine derivative substrates with rigid and bulky substituents at the *para* position. However, ESI-MS data for compound **3** also show a small side peak corresponding to the incorporation of phenylalanine, a result we have observed previously. The remaining satellite peaks for these compounds correspond to common metal adducts in ESI-MS. Additionally, results obtained for **1** demonstrated that both *meta* and *para* positions can be occupied without detriment to expression levels. These

results, coupled with our previous endeavors, led us to wonder if phenylalanine derivatives with long-chain *meta*-substituents could serve as substrates of PylRS(N346A/C348A) for genetic incorporation as well. To investigate this hypothesis, a series of *meta*-alkoxy and *meta*-acyl phenylalanines with substituent chain lengths of up to six carbons were synthesized. We chose these specific derivatives because the parent NCAAs *meta*-methoxy-phenylalanine and *meta*-acetylphenylalanine act as efficient substrates for PylRS(N346A/C348A). The synthesis of *meta*-alkoxyphenylalanines was straightforward, starting with a published route to obtain protected *meta*-tyrosine, at which point the intermediate was subjected to various alkyl halides to afford different derivatives. Acidic deprotection then yielded free amino acids as racemic chloride salts. The synthesis of *meta*-acylphenylalanines was more divergent. Alkyl Grignards were added to a solution of *meta*-tolunitrile, which afforded acylbenzenes upon acidic workup. Radical bromination and then displacement with diethylacetamidomalonate afforded protected *meta*-acylphenylalanines that were deprotected in 6 M HCl to obtain free amino acids. More detailed synthetic routes can be found later in this Section.

With the desired NCAAs in hand, we thenceforth tested their incorporation efficacies at amber codons using the PylRS(N346A/C48A)– $tRNA^{Pyl/CUA}$ pair. The *E. coli* cells used for these compounds harbored two plasmids, pEVOL-pylT-PylRSN346A/C348A and pET-pylT-sfGFPS2TAG’. pET-pylT-sfGFPS2TAG’ contains a sequence-optimized sfGFP with an amber mutation at its S2 position (sfGFPS2TAG’). In comparison to the sfGFP2TAG gene in pET-pylT-sfGFP2TAG, sfGFPS2TAG’ has one

more alanine residue in front of the amber mutation. Growing this cell in minimal media without NCAA yielded a minimal expression level of full-length sfGFP. However, all ether NCAAs **6–10** (2 mM) in the medium promoted the synthesis of sfGFP with a designated NCAA incorporated (Figure **2.2.A**). In comparison to phenylalanine derivatives with small *meta*-substituents such as **5**, **6–10** apparently have low incorporation levels. Molecular weights of purified sfGFP variants determined by ESI-MS agreed well with the theoretical values corresponding to a designated NCAA at the S2 position and the first methionine hydrolysed (Figure **2.2.B**). The removal of the first methionine is due to the insertion of alanine after it. A number of smaller signals can be observed, but they largely correspond to common metal adducts; the expected masses were always the major signal. Compounds **8**, **9**, and **10** have low solubility; when added to the medium at 2 mM, compound **10** was observed to precipitate after 12 h of expression. The low sfGFP expression levels for **8**, **9**, and **10** may be partially due to the toxicity of the compounds; indeed, smaller pellet sizes are observed for **8** and **9**. Although the sfGFP expression level for **9** was very low, the purified sfGFP displayed an ESI-MS molecular weight that still matched the theoretical value of sfGFP with **9** incorporated at S2, indicating that a low concentration of **9** was still sufficient to observe incorporation of **9** at the amber mutation site.

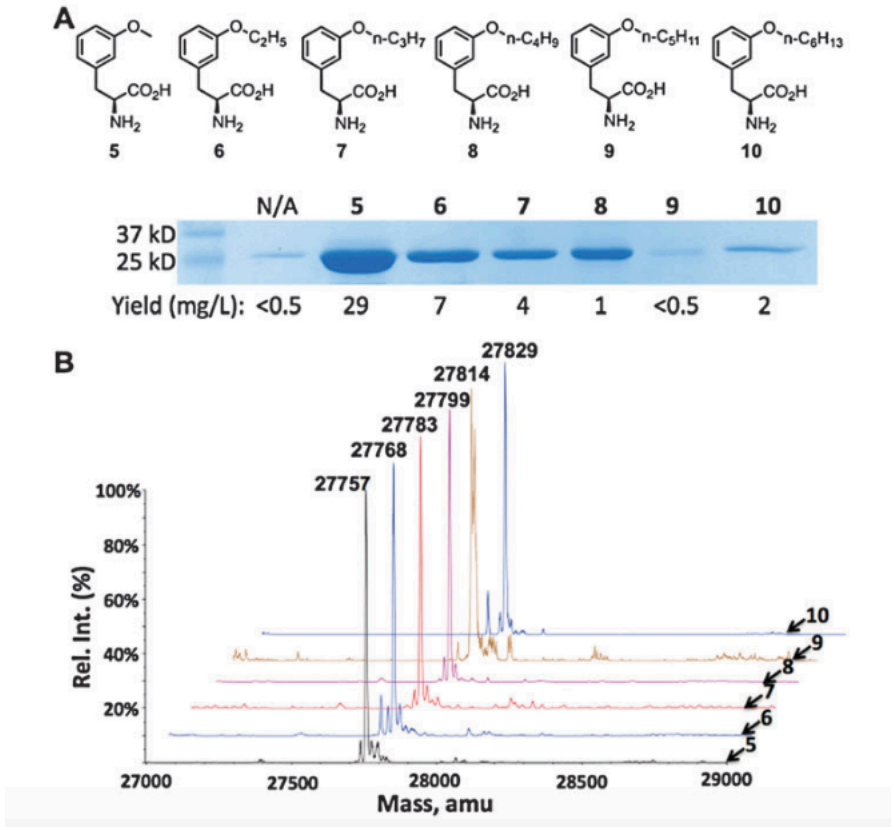


Figure 2.2. (A) Structures of **5–10** and their site-specific incorporation into sfGFP at its S2 position. (B) Deconvoluted ESI-MS spectra of sfGFP variants incorporated with **5–10**. Their theoretical values are 27 758 Da for **5**, 27 772 Da for **6**, 27786 Da for **7**, 27800 Da for **8**, 27814 Da for **9**, and 27828 Da for **10**.

Overall, addition of ketone derivatives **12–15** at 2 mM to the medium promoted high sfGFP expression yields, and longer alkyl lengths had less of an impact on protein yields in comparison to the ether series **6–10**, though the sfGFP expression levels for **12–15** are lower than that for **11** (Figure **2.3.A**). This series of NCAAs are also readily soluble, with no precipitation observed in the medium after overnight incubation. ESI-

MS analysis of the purified sfGFP variants confirmed high incorporation fidelities of **12–15** at the S2 site.

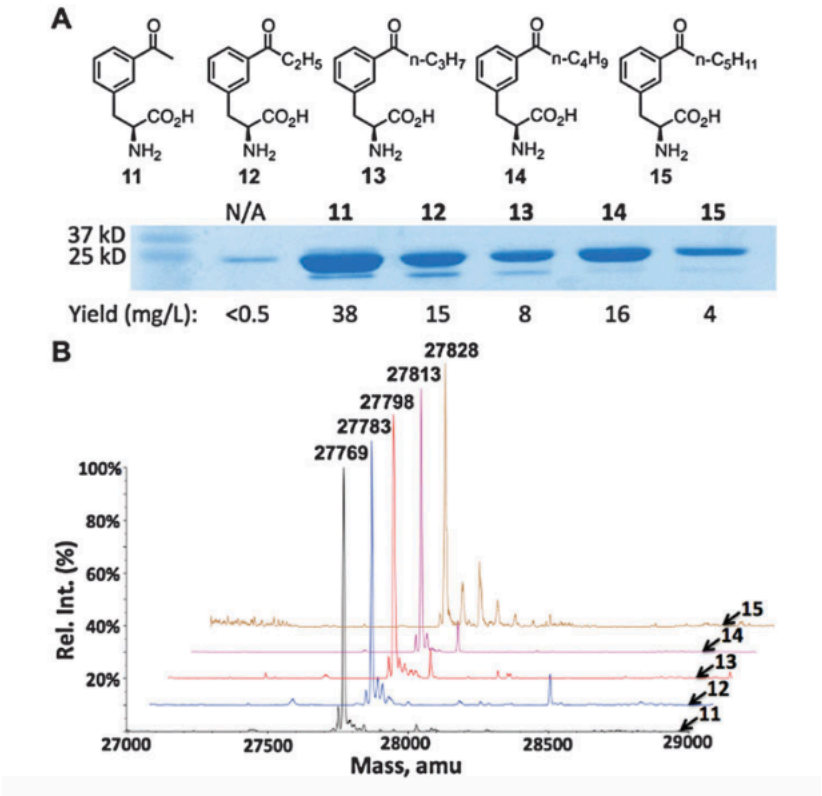


Figure 2.3. (A) Structures of **11–15** and their site-specific incorporation into sfGFP at its S2 position. (B) Deconvoluted ESI-MS spectra of sfGFP variants incorporated with **11–15**. Their theoretical values are 27770 Da for **11**, 27784 Da for **12**, 27798 Da for **13**, 27812 Da for **14**, and 27826 Da for **15**. Compounds **13** and **14** show small signals corresponding to an N-terminal methionine on sfGFP. Compound **15** has several small signals attributed to sodium and potassium adducts.

Among all of the novel NCAAs that can be taken by PylRS(N346A/ C348A), **2** has an active azide functionality for a click reaction with an alkyne and **12–15** contain a ketone group that potentially reacts with a hydroxylamine. Both functionalities can be applied for site-selective labeling of proteins incorporated with **2** and **12–15**. Since labeling of sfGFP incorporated with **11** with a hydroxylamine dye was demonstrated previously,[113] we chose to demonstrate the selective labeling of **2** using a diarylcyclooctyne dye **D1** in this study (Figure **2.4**). **D1** contains a strained alkyne that undergoes a spontaneous reaction with an azide.[114] Incubating sfGFP incorporated with **2** with **D1** overnight led to an intensely fluorescently labeled protein; however, the same reaction with sfGFP incorporated with **3** did not yield any fluorescently labeled final product. This result indicates that genetically incorporated **2** can be applied to site-specifically introduce biophysical and biochemical probes to proteins for a large variety of studies.

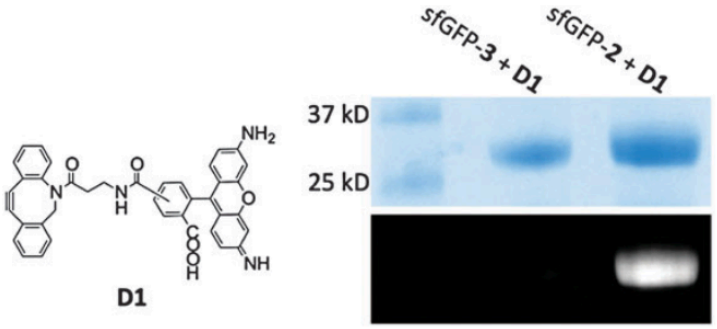


Figure 2.4. Labeling of sfGFP incorporated with **2** (sfGFP-**2**) and sfGFP incorporated with **3** (sfGFP-**3**) with dye **D1**. The top panel shows the Coomassie blue stained SDS-PAGE gel and the bottom panel shows the fluorescent image of the same gel under UV irradiation before the gel was stained with Coomassie blue.

2.3. Conclusion

In summary, we have shown that thirteen novel NCAAs were genetically incorporated into protein at the amber codon in *E. coli* using the PylRS(N346A/C348A)–$tRNA^{Pyl/CUA}$ pair. This result, coupled with our previous findings, shows a surprisingly broad substrate scope for PylRS(N346A/C348A). Investigations are underway to determine aspects of the active site pocket of PylRS(N346A/C348A) that lead to this broad substrate spectrum. The current study has great implications in understanding amino acid structure tolerance of the protein translation system. The expanded genetically encoded NCAA pool can also be applied to generate phage and *E. coli* displayed peptide libraries with expanded chemical moieties for drug discovery, a direction we are actively pursuing at the current stage.

This work was supported in part by the Welch Foundation (grant A-1715), the National Science Foundation (grant CHEM-1148684), and the National Institute of Health (grant 1R01CA161158).

2.4. Detailed Experimental Protocols

2.4.1. Superfolder Green Fluorescent Protein Expression

The constructs used to incorporate compounds **1**-**4** in this study and corresponding protein purification and characterization are identical to previous reports.[115,116] For compounds **5**-**15**, BL21(DE3) *E. coli* cells containing pEVOL-PylT-pylRS-PylRSN346A/C348A (Cm^r) and pET-PylT-sfGFP-S2TAG' (Amp^r) vectors were grown in 500 mL of LB media with ampicillin (100 μg/mL) and chloramphenicol (34 μg/mL) until the OD_{600} reached 1.0-1.3, at which point the cells were pelleted at 4,000

r.p.m. for 20 min, then washed and resuspended in 30 mL H_2O. The resuspended cells were added as 5 mL aliquots to 45 mL of minimal media (33.7 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 9.4 mM NH_4Cl, 1 mM $MgSO_4$, 0.3 mM $CaCl_2$, 1% glycerol) supplemented with 2 mM non-canonical amino acid (NAA), 1 mM IPTG, and 0.2% arabinose, at which point protein expression was allowed to occur for 12 h. The cells were then pelleted at 4,000 r.p.m. for 20 min, resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8) and lysed via sonication. The crude lysate was then centrifuged at 10,000 r.p.m. for 1 hour, and the supernatant was treated with imidazole to a final concentration of 10 mM. Next, the supernatant was subsequently incubated with Ni^{2+}-NTA resin for 1 hour at 4 ºC. The resin was washed with lysis buffer containing 10 mM imidazole (3x column volume) and 20 mM imidazole (3x column volume), then eluted with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 500 mM imidazole, pH 8). The protein was dialyzed against 10 mM Tris buffer, and if necessary, the proteins were concentrated using Amicon Ultracel-10k centrifugal filter units. Purity of the proteins was confirmed via 15% SDS-PAGE and ESI-MS analysis. Yields were determined using a commercially available BCA protein assay kit (Thermo Scientific).

2.4.2. sfGFP-2 Protein Labeling

To 16 μl of 1X Phosphate buffered saline (PBS) was added 2 μl sfGFP-**2** (3 mg/mL, 1 μM) and 2 μl Dibenzylcyclooctyne (DBCO) dye (25 mM, DMSO stock solution)[117] and was allowed to react at room temperature overnight. The protein was then precipitated with 180 μl of methanol and left at -20 ºC for 1 h. Next, the precipitated protein was centrifuged for 5 min, 14,000 r.p.m. and washed twice with

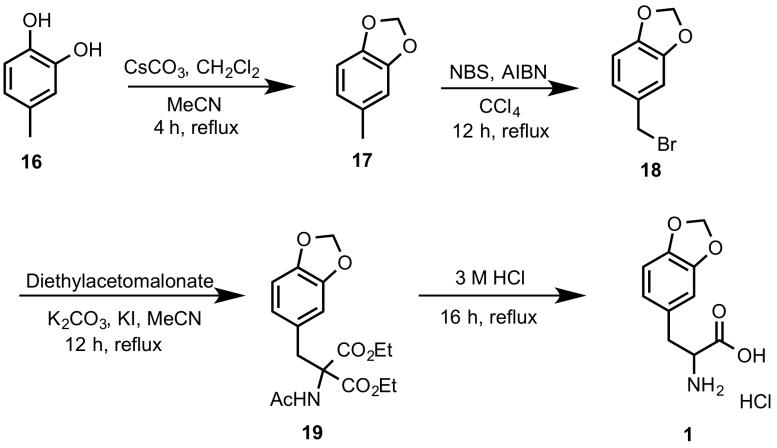
100% methanol. Finally, the residue was resuspended in 20 μl H_2O and subjected to SDS-PAGE analysis. The control experiment was identical except 4 μl sfGFP-**3** (5 mg/mL, 0.9 μM) was added to 14 μl PBS, followed by 2 μl DBCO dye.

2.4.3. ESI-MS Analysis of Intact Proteins

Nanoelectrospray ionization in positive mode was performed using an Applied Biosystems QSTAR Pulsar (Concord, ON, Canada) equipped with a nanoelectrospray ion source. Solution was flowed at 700 nL/min through a 50 μm ID fused-silica capillary that was tapered at the tip. Electrospray needle voltage was held at 2100 V.

2.4.4. Organic Synthesis

Reactions were carried out using oven-dried glassware and under an atmosphere of argon, where appropriate. Reagents were purchased and used without further purification. NMR spectra were obtained with Inova 300 and Mercury 300 MHz instruments.



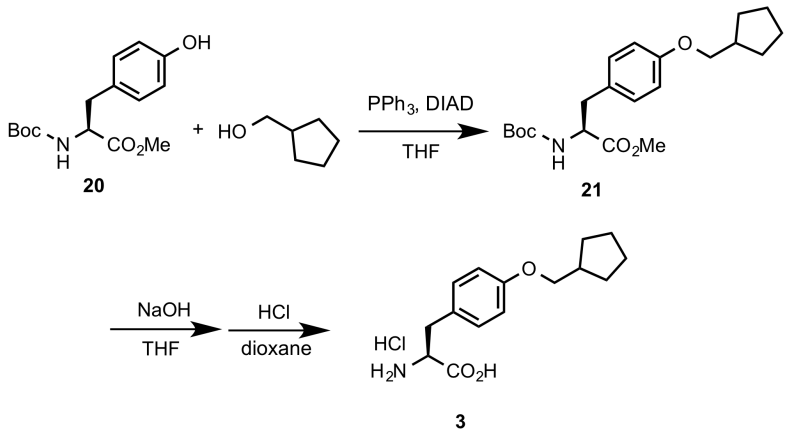
*2-Amino-3-(benzo[d][1,3]dioxol-5-yl)propanoic Acid Hydrochloride (**1**)*

To a solution of catechol **16** (5 g, 8.06 mmol) in MeCN (25 mL) was added Cs_2CO_3 (3.93g 12.09 mmol), followed by CH_2Cl_2 (5.21 mL, 8.06 mmol). The resulting mixture was refluxed for 4 hours. After cooling to room temperature, the reaction mixture was concentrated and applied to column chromatography, yielding the acetal **17** in 40% yield (2.12 g).[118,119] ^{1}HNMR (300 MHz, $CDCl_3$) δ 6.71(m, 3H), 5.93(s, 2H), 2.27(s, 3H).

The compound **17** (2.0 g, 14.7 mmol) was dissolved in CCl_4 (15 mL), followed by addition of NBS (3.12 g, 17.62 mmol) and AIBN (0.337 g, 2.05 mmol). The mixture was heated to reflux for 12 h under the protection of argon. After cooling, the precipitate was removed via filtration and the filtrate was concentrated and dried under vacuum to afford the known compound **18**,[120] which was used directly in the next step without further purification.

To a solution of compound **18** (3.5 g, 16.27 mmol) in anhydrous acetonitrile (15 mL) was added diethyl 2-acetamidomalonate (3.88 g, 17.88 mmol), K_2CO_3 (4.49 g, 32.53 mmol) and KI (1.0 g, 15.38 mmol). The resulting mixture was heated to reflux for 12 h under the protection of argon, then cooled to room temperature. The solid was filtered, the solvent was removed under reduced pressure, and the residue was purified via column chromatograph with hexanes/ethyl acetate (3:1 v/v) as eluent to give the pure product **19** (60% yield, 3.42 g). ^{1}HNMR (300 MHz, $CDCl_3$) δ 6.69 (d, J = 8.2Hz, 1 H), 6.57(d, J = 7.9 Hz, 1 H), 6.47 (s, 1H), 5.92 (s, 2H), 4.30-4.22 (m, 4H), 3.56 (s, 2H), 2.04 (s, 3H), 1.32-1.23 (m, 6H).

A suspension of compound **19** (3.0 g, 8.54 mmol) in 3 M HCl (91.16 mL, 32 eq.) was heated to reflux for 16 h before cooling to room temperature. Water was evaporated under reduced pressure and the solid was collected. The solid was washed with Et_2O (10 mL × 3) and then dried under vacuum to afford compound **1** as a solid (68% yield, 1.42 g). ^{1}H NMR (300 MHz, CD_3OD) δ 6.78-6.66 (m, 3H), 5.85 (s, 2H), 4.14 (t, J = 6.6 Hz, 1 H), 3.09 (ABq, J = 15.3, 6.6 Hz, 2 H). ^{13}C NMR (75 MHz, CD_3OD) δ 169.8, 148.1, 147.3, 127.6, 122.6, 109.1, 108.2, 101.1, 53.9, 39.5.

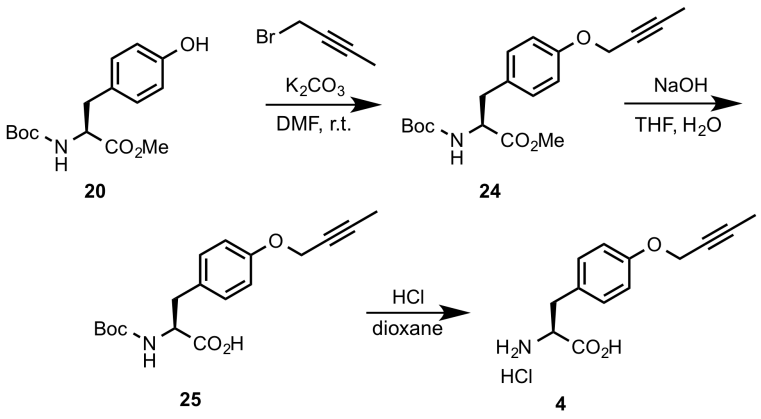


*(S)-2-Amino-3-(4-(cyclopentylmethoxy)phenyl)propanoic Acid Hydrochloride (**3**)*

N-Boc-*o*-methyl-L-tyrosine (**20)** (1.3 g, 4.4 mmol), cyclopentylmethanol (0.57 mL, 5.28 mmol), and triphenylphosphine (1.73 g, 6.6 mmol) were added to a round bottom flask, then the atmosphere evacuated and replaced with argon. THF (10 mL) was then added and the reaction mixture was cooled to 0 ºC, at which point DEAD (1.3 mL,

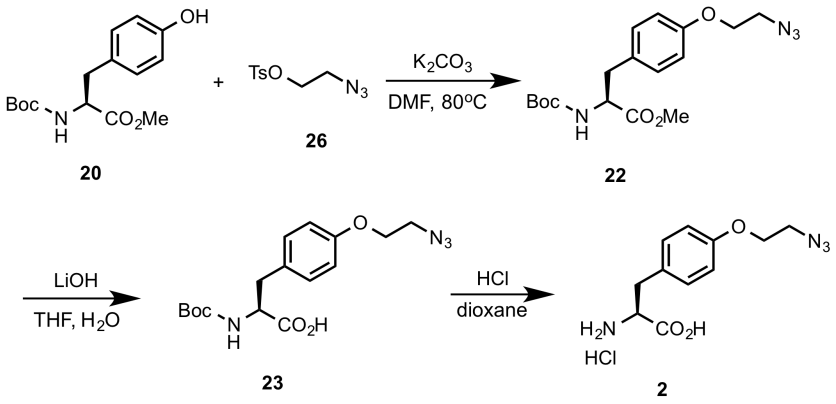
6.6 mmol) was added dropwise and stirred overnight. Upon completion, the reaction mixture was concentrated and purified via column chromatography (gradient elution, 25% to 50% ethyl acetate/hexanes) to give the product **21** in 78% yield (1.3 g).

This compound was deprotected using the previously described procedure;[115] treatment with 5 mL 1 M NaOH/THF for two hours, followed by Boc deprotection in 5 mL 4 M HCl/dioxane to give the free amino **3** acid in 97% yield (898 mg), two steps. ^{1}H NMR (300 MHz, CD_3OD) δ 7.16 (d, J = 8.4 Hz, 2 H), 6.87 (d, J = 8.7, 6.9 Hz, 2 H), 4.15 (t, J = 5.4 Hz, 1 H), 3.80 (d, J = 6.9 Hz, 2 H), 3.14 (dd, J = 7.8, 5.4 Hz, 2 H), 2.31 (m, 1 H), 1.82 (m, 2 H), 1.61 (m, 4 H), 1.36 (m, 2 H). ^{13}C NMR (75 MHz, CD_3OD) δ 169.9, 158.9, 130.0, 125.6, 114.6, 71.8, 53.8, 38.9, 35.0, 28.9, 24.9.



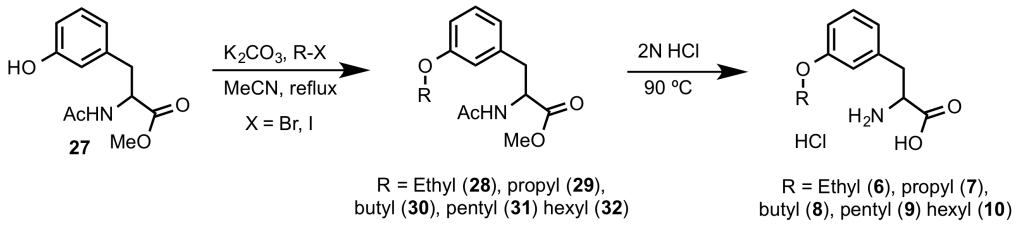
*(S)-2-Amino-3-(4-(but-2-yn-1-yloxy)phenyl)propanoic Acid Hydrochloride (**4**)*

N-Boc-*o*-methyl-L-tyrosine (1 g, 3.39 mmol) was treated with 3-bromo-2-propyne (0.36 mL, 4.06 mmol) and potassium carbonate (1.4 g, 10.2 mmol) in DMF and left to react overnight at room temperature. The reaction mixture was then diluted with ethyl acetate and washed with 3 M HCl. The organic layers were then dried, concentrated, and purified via flash chromatography (gradient elution, 25% to 50% ethyl acetate/hexanes) to afford the protected azide **24** in 30% yield (370 mg). Deprotection using the protocol listed above gave the title compound **4** in 44% yield (274 mg). ^{1}H NMR (300 MHz, CD_3OD) δ 7.10 (d, J = 8.7 Hz, 2 H), 6.83 (d, J = 8.7 Hz, 2 H), 4.53 (m, 2 H), 4.08 (q, J = 7.5 Hz, 1 H), 3.02 (dd, J = 14.7, 7.5 Hz, 2 H), 1.68 (s, 3 H). ^{13}C NMR (75 MHz, CD_3OD) δ 168.1, 155.8, 128.4, 124.6, 113.3, 81.0, 72.0, 53.9, 52.0, 33.2.



*(S)-2-Amino-3-(4-(2-azidoethoxy)phenyl)propanoic Acid Hydrochloride (**2**)*

N-Boc-*o*-methyl-L-tyrosine **20** (1.8 g, 6.09 mmol) was treated with compound **26** (1.8 g, 7.31 mmol) and potassium carbonate (3.3 g, 24.4 mmol) in DMF (10 mL) and the reaction was heated to 80 ºC. When the reaction was complete based on TLC, the reaction mixture was worked up as usual and purified via flash chromatography (gradient elution, 25% to 50% ethyl acetate/hexanes) to yield compound **22**. Deprotection was achieved using 20 mL 1 M LiOH, followed by treatment with 5 mL 4 M HCl/Dioxane to afford free amino acid **2** (50%, two steps, 873 mg). ^{1}H NMR (300 MHz, D_2O) δ 7.16 (d, *J* = 8.4 Hz, 2 H), 6.88 (d, *J* = 8.4 Hz, 2 H), 4.11 (m, 3 H), 3.50 (m, 2 H), 3.11 (m, *J* = 7.2, 6.9, 5.4 Hz, 2 H). ^{13}C NMR (75 MHz, D_2O) δ 169.8, 158.1, 130.2, 126.4, 114.7, 67.0, 53.8, 49.8, 35.0.



R = Ethyl (**28**), propyl (**29**), butyl (**30**), pentyl (**31**) hexyl (**32**)

R = Ethyl (**6**), propyl (**7**), butyl (**8**), pentyl (**9**) hexyl (**10**)

Representative Procedure for m-Tyrosine Alkylation

*Methyl 2-Acetamido-3-(3-ethoxyphenyl)propanoate (**28**)*

To a solution of *m*-tyrosine[121] **27** (1.5 g, 6.32 mmol) in DMF (12.64 mL) was added K_2CO_3 (2.84 g, 20.55 mmol), followed by ethyl iodide (0.76 mL, 9.48 mmol). The

reaction was stirred at room temperature for 18 h, then quenched with 16 mL H_2O and 79 mL ethyl acetate. The organic layer was extracted three times with 40 mL H_2O, then dried with sodium sulfate and concentrated. Column chromatography (gradient elution, 50 to 75% ethyl acetate/hexanes) afforded the compound as a yellow, crystalline solid (72% yield, 1.2 g). ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.17 (t, *J* = 7.8 Hz, 1H), 6.76 (dd, *J* = 8.1, 5.7 Hz, 1 H), 6.66 (m, 2 H), 6.04 (d, *J* = 7.5 Hz, 1 H), 4.85 (dt, *J* = 7.8, 5.7 Hz, 1 H), 3.97 (q, *J* = 6.9 Hz, 2 H), 3.71 (s, 3 H), 3.07 (m, *J* = 8.1, 5.7 Hz, 2 H), 1.97 (s, 3 H), 1.38 (t, *J* = 6.9 Hz, 3 H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 172.2, 169.7, 159.2, 137.3, 129.6, 121.5, 115.6, 113.2, 63.4, 53.1, 52.4, 37.9, 23.3, 14.9.

*Methyl 2-Acetamido-3-(3-propoxyphenyl)propanoate (**29**)*

Synthesized according to the general procedure with *m*-tyrosine (1.5 g, 6.32 mmol), DMF (12.64 mL), K_2CO_3 (2.84g, 20.55 mmol), and n-propyl bromide (0.86 mL, 9.48 mmol) to yield a yellow, crystalline solid, (68% yield, 1.2 g). ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.18 (t, *J* = 8.1 Hz, 1 H), 6.79 (m, *J* = 5.7 Hz, 1 H), 6.66 (m, 2H), 5.88 (d, *J* = 7.8 Hz, 1 H), 4.87 (dt, *J* = 7.8, 5.7 Hz, 1H), 3.88 (t, *J* = 6.6 Hz, 2 H), 3.74 (s, 3 H), 3.09 (m, *J* = 8.1, 5.7 Hz, 2 H), 1.99 (s, 3H), 1.79 (sx, *J* = 7.2 Hz, 2 H), 1.03 (t, *J* = 7.2 Hz, 3 H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 172.1, 169.7, 159.3, 137.3, 129.6, 121.4, 115.6, 113.2, 69.5, 53.1, 52.4, 37.9, 23.2, 22.6, 10.6.

*Methyl 2-Acetamido-3-(3-butoxyphenyl)propanoate (**30**)*

Synthesized according to the general procedure with *m*-tyrosine (1.0 g, 4.21 mmol), DMF (8.42 mL), K_2CO_3 (1.89g, 13.69 mmol), and n-butyl iodide (0.72 mL, 6.32 mmol) to yield a yellow, crystalline solid (65% yield, 0.8 g). ^{1}H NMR (300 MHz,

$CDCl_3$) δ 7.19 (t, *J* = 7.8 Hz, 1 H), 6.77 (d, *J* = 7.2 Hz, 1 H), 6.64 (m, 2 H), 5.87 (d, *J* = 6.9 Hz, 1 H), 4.87 (dt, *J* = 7.8, 5.4 Hz, 1 H), 3.92 (t, *J* = 6.6 Hz, 2 H), 3.74 (s, 3 H), 3.09 (dd, *J* = 8.1, 5.7 Hz, 2 H), 1.99 (s, 3 H), 1.76 (p, *J* = 6.9, 6.6, 6.3 Hz, 2 H), 1.46 (m, 2 H), 0.97 (t, *J* = 7.2 Hz, 3 H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 172.2, 169.7, 159.4, 137.3, 129.6, 121.4, 115.6, 113.2, 67.7, 53.1, 52.4, 37.9, 31.4, 23.2, 19.3, 13.9.

*Methyl 2-Acetamido-3-(3-(pentyloxy)phenyl)propanoate (**31**)*

Synthesized according to the general procedure with *m*-tyrosine (1.5 g, 6.32 mmol), DMF (12.64 mL), K_2CO_3 (2.84g, 20.55 mmol), and n-pentyl bromide (1.18 mL, 9.48 mmol) to yield a yellow, crystalline solid (72% yield, 1.4 g). ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.19 (t, *J* = 8.1 Hz, 1 H), 6.78 (m, 1 H), 6.63 (m, 2H), 5.87 (d, *J* = 6.9 Hz, 1 H), 4.82 (dt, *J* = 7.8, 5.7 Hz, 1H), 3.91 (t, *J* = 6.6 Hz, 2 H), 3.74 (s, 3 H), 3.10 (dq, *J* = 8.1, 6.0 Hz, 2 H), 1.99 (s, 3H), 1.77 (p, *J* = 7.2, 6.9, 6.6 Hz, 2 H), 1.39 (m, 4 H), 0.93 (t, *J* = 7.2 Hz, 3 H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 172.1, 169.7, 159.2, 137.3, 129.5, 121.2, 115.5, 113.0, 67.8, 53.1, 52.3, 37.8, 28.9, 28.2, 23.0, 22.4, 14.0.

*Methyl 2-Acetamido-3-(3-(hexyloxy)phenyl)propanoate (**32**)*

Synthesized according to the general procedure with *m*-tyrosine (1.0 g, 4.21 mmol), DMF (8.42 mL), K_2CO_3 (1.89g, 13.69 mmol), and 1-iodohexane (0.93 mL, 6.32 mmol) to yield a yellow, crystalline solid (44% yield, 0.6 g). ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.18 (t, *J* = 8.1 Hz, 1 H), 6.77 (dd, *J* = 8.1, 5.7 Hz, 1 H), 6.63 (m, 2H), 5.88 (d, *J* = 7.2 Hz, 1 H), 4.87 (dt, *J* = 7.8, 5.7 Hz, 1H), 3.91 (t, *J* = 6.6 Hz, 1 H), 3.74 (s, 3 H), 3.09 (m, *J* = 8.1, 5.7 Hz, 2 H), 1.99 (s, 3H), 1.77 (p, *J* = 8.1, 6.6 Hz, 2 H), 1.44 (m, 4 H), 1.33 (m, 4H), 0.91 (t, *J* = 7.2 Hz, 3 H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 172.2, 169.7,

159.4, 137.3, 129.6, 121.4, 115.6, 113.2, 68.0, 53.1, 52.4, 37.9, 31.7, 29.3, 25.8, 23.3, 22.7, 14.1.

Representative Procedure for meta-alkoxy Phenylalanine Deprotection

*2-Amino-3-(3-ethoxyphenyl)propanoic Acid Hydrochloride (**6**)*

A round bottomed flask charged with compound **28** (1.1 g, 4.15 mmol) was treated with 14.3 mL 2 M HCl, and the resulting suspension was heated to 100 ºC, which was refluxed overnight until complete consumption of starting material as observed on TLC. The reaction mixture was then cooled to room temperature and concentrated *in vacuo*. If necessary, the compound was resubjected to the reaction conditions due to the persistence of methyl ester and/or N-acyl amide signals in ^{1}H and ^{13}C NMR. The title compound was obtained as a white solid in 45% yield (463 mg). ^{1}H NMR (300 MHz, D_2O) δ 7.35 (t, J = 7.8 Hz, 1 H), 6.93 (m, 3 H), 4.31 (t, J = 7.5 Hz 1 H), 4.11 (q, J = 6.9 Hz, 2 H), 3.24 (dq, J = 14.7, 5.7 Hz, 2 H), 1.37 (t, J = 6.9 Hz, 3 H). ^{13}C NMR (75 MHz, D_2O) δ 171.4, 158.3, 135.6, 130.4, 122.1, 115.5, 114.1, 64.2, 54.0, 35.5, 13.8.

*2-Amino-3-(3-propoxyphenyl)propanoic Acid Hydrochloride (**7**)*

Prepared according to the general procedure using compound **29** (1.0 g, 3.58 mmol) and 12.34 mL 2 M HCl. Obtained as a white solid in 43% yield (402 mg). ^{1}H NMR (300 MHz, D_2O) δ 7.34 (t, J = 7.5 Hz, 1 H), 6.90 (m, 3 H), 4.29 (t, J = 5.4 Hz, 1 H), 3.97 (t, J = 6.6 Hz), 3.21 (dd, J = 14.7, 5.7 Hz, 2 H), 1.73 (sx, J = 7.5, 7.2, 6.9, 6.6 Hz, 2 H), 0.96 (t, J = 7.5, 3 H). ^{13}C NMR (75 MHz, D_2O) δ 171.3, 158.5, 135.5, 130.3, 122.0, 115.5, 114.1, 70.1, 53.9, 35.4, 21.7, 9.5.

*2-Amino-3-(3-butoxyphenyl)propanoic Acid Hydrochloride (**8**)*

Prepared according to the general procedure using compound **30** (0.8 g, 2.73 mmol) in 9.4 mL 2 M HCl to obtain the title compound in 30% yield (222 mg). ^{1}H NMR (300 MHz, DMSO) δ 8.53 (s, 2 H), 7.21 (t, *J* = 7.8 Hz, 1 H), 6.89 (s, 1 H), 6.82 (d, *J* = 7.8 Hz, 2 H), 4.14 (t, *J* = 5.4 Hz, 1 H), 3.94 (t, *J* = 6.6 Hz, 2 H), 3.12 (d, *J* = 6.0 Hz, 2 H), 1.69 (p, *J* = 7.5, 6.9, 6.6 Hz, 2 H), 1.45 (sx, *J* = 7.5, 7.2 Hz, 2 H), 0.93 (t, *J* = 7.2 Hz, 3 H). ^{13}C NMR (75 MHz, DMSO) δ 170.2, 158.7, 136.4, 129.5, 121.5, 115.6, 113.1, 66.9, 53.1, 35.5, 30.8, 18.8, 13.7.

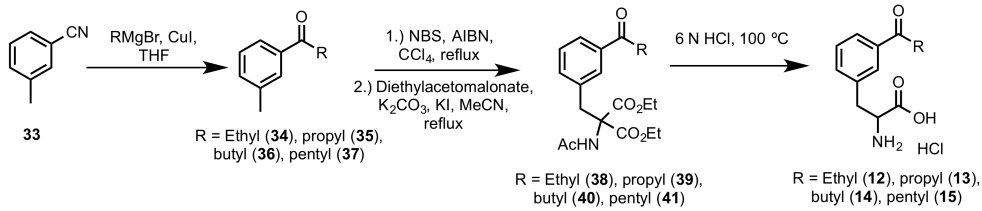
*2-Amino-3-(3-(pentyloxy)phenyl)propanoic Acid Hydrochloride (**9**)*

Prepared according to the general procedure using compound **31** (1.0 g, 3.25 mmol) in 11.2 mL 2 M HCl to obtain the title compound in 61% yield (570 mg). ^{1}H NMR (300 MHz, DMSO) δ 8.49 (s, 2 H), 7.21 (t, *J* = 7.5 Hz, 1 H), 6.88 (s, 1 H), 6.82 (d, *J* = 8.1 Hz, 2 H), 4.14 (t, *J* = 6.0 Hz, 1 H), 3.94 (t, *J* = 6.6 Hz, 1 H), 3.11 (d, *J* = 6.0 Hz, 2 H), 1.71 (p, *J* = 6.9, 6.6, 6.3 Hz, 2 H), 1.36 (m, 4 H), 0.89 (t, *J* = 6.6 Hz, 3 H). ^{13}C NMR (75 MHz, DMSO) δ 170.2, 158.7, 136.4, 129.5, 121.6, 115.6, 113.1, 67.2, 53.0, 35.6, 28.4, 27.7, 21.9, 13.9.

*2-Amino-3-(3-(hexyloxy)phenyl)propanoic Acid Hydrochloride (**10**)*

Prepared according to the general procedure using compound **32** (0.6 g, 1.87 mmol) in 6.44 mL 2 M HCl to yield the compound as a white solid in 65% yield (367 mg). ^{1}H NMR (300 MHz, DMSO) δ 8.48 (s, 3 H), 7.21 (t, *J* = 7.8 Hz, 1 H), 6.88 (s, 1 H), 6.82 (d, *J* = 7.8 Hz, 2 H), 4.13 (m, *J* = 5.1, 4.5 Hz, 1 H), 3.93 (t, *J* = 6.6 Hz, 2 H), 3.10 (d, *J* = 6.3 Hz, 2 H), 1.69 (p, *J* = 7.8, 6.6 Hz, 2 H), 1.41 (m, 2 H), 1.29 (m, 4 H), 0.88 (t,

J = 6.6 Hz, 3 H). ^{13}C NMR (75 MHz, DMSO) δ 170.2, 158.7, 136.3, 129.5, 121.5, 115.6, 113.1, 67.2, 53.0, 35.6, 31.0, 28.7, 25.2, 22.1, 13.9.



33

R = Ethyl (**34**), propyl (**35**), butyl (**36**), pentyl (**37**)

R = Ethyl (**38**), propyl (**39**), butyl (**40**), pentyl (**41**)

R = Ethyl (**12**), propyl (**13**), butyl (**14**), pentyl (**15**)

Representative Procedure for Grignard Addition to m-Tolunitrile

*1-(m-Tolyl)propan-1-one (**34**)*

Ethyl magnesium bromide (1 M/THF, 85.36 mL, 85.36 mmol) was added dropwise to a solution of *m*-tolunitrile (10.25 mL, 85.36 mmol) and copper (I) iodide (40.64 mg, 0.213 mmol) in anhydrous THF (170.72 mL) under argon. After stirring for 22 h, the reaction was quenched with approx. 5 mL 1 M HCl at 0 ºC and stirred for 4 h, allowing the reaction to warm to room temperature. The resulting layers were separated and the organic layer was dried with $MgSO_4$, filtered, and concentrated. The crude, yellow oil was purified via silica gel chromatography (gradient, 0 to 10% EtOAc/Hex) to afford the known compound[122] **34** in 94% yield (9.4 g) as a yellow oil. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.76 (m, 1H), 7.38 (m, 3H), 2.99 (q, J = 7.2 Hz, 2H), 2.41 (s, 3H), 1.22 (t, J = 7.2 Hz, 3H).

*1-(m-Tolyl)butan-1-one (**35**)*

Prepared according to the general procedure with Propylmagnesium bromide (2 M/THF, 42.68 mL, 85.36 mmol), *m*-tolunitrile (10.25 mL, 85.36 mmol), and CuI (40.64 mg, 0.213 mmol) to yield the known compound[123] in 96 % yield (13.24 g). ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.77 (s, 2H), 7.36 (s, 2H), 2.94 (t, J = 7.2 Hz, 2H), 2.41 (s 3H), 1.76 (sx, J = 7.2 Hz, 2H), 1.00 (t, J = 7.2 Hz, 3H).

*1-(m-Tolyl)pentan-1-one (**36**)*

Prepared according to the general procedure with butylmagnesium chloride (2 M/THF, 42.68 mL, 85.36 mmol), *m*-tolunitrile (10.25 mL, 85.36 mmol), and CuI (40.64 mg, 0.213 mmol) to yield the known compound[124] in 76% yield (11.43 g). ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.75 (m, 2H), 7.34 (m, 2H), 2.95 (t, J = 7.2 Hz, 2H), 2.41 (s, 3H), 1.71 (p, J = 7.5 Hz, 2H), 1.41 (sx, J = 7.2 Hz, 2H), 0.95 (t, J = 7.5 Hz, 3H).

*1-(m-Tolyl)hexan-1-one (**37**)*

Prepared according to the general procedure with Pentylmagnesium bromide (1 M/THF, 42.68 mL, 85.36 mmol), *m*-tolunitrile (10.25 mL, 85.36 mmol), and CuI (40.64 mg, 0.213 mmol) to yield the known compound[125] in 94% yield (15.3 g). ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.74 (m, 1H), 7.46 (s, 1H), 7.37 (m, 2H), 2.94 (t, J = 6.9 Hz, 2H), 2.41 (s, 3H), 1.73 (p, J = 7.2 Hz, 2H), 1.35 (m, 4H), 0.91 (t, J = 6.6 Hz, 3H).

Representative Procedure for the Synthesis of Protected Ketones

*Diethyl 2-Acetamido-2-(3-propionylbenzyl)malonate (**38**)*

To a solution of *m*-Keto toluene **34** (8.0 g, 53.98 mmol) in CCl_4 (134.95 mL) was added NBS (10.57 g, 59.38 mmol) and AIBN (2.67 g, 16.19 mmol), and the resulting suspension was refluxed overnight. Upon completion, the reaction was filtered and concentrated, and the resulting crude material was subjected to the next step without further purification.

A round-bottom flask was charged with brominated **34** (4.24 g, 19.09 mmol), diethyl 2-acetamidomalonate (3.73 g, 17.18 mmol), K_2CO_3 (5.28 g, 38.18 mmol), and KI (3.17 g, 19.09 mmol), followed by 119.3 mL of MeCN, and the resulting suspension was heated to reflux and stirred overnight. Upon completion, the reaction was cooled to room temperature, filtered with celite, and concentrated. Purification via silica gel chromatography (gradient elution, 0 to 30% EtOAc/Hex) afforded **38** as a yellow solid in 58% yield (3.6 g). ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.83 (d, J = 7.8 Hz, 1 H), 7.64 (s, 1 H), 7.36 (t, J = 7.8, 7.5 Hz, 1 H), 7.20 (d, J = 8.4 Hz, 1 H), 6.52 (s, 1 H), 4.28 (q, J = 7.2, 6.9 Hz, 4 H), 2.96 (q, J = 7.2 Hz, 2 H), 2.05 (s, 3 H), 1.31 (t, J = 7.2 Hz, 6 H), 1.21 (t, J = 7.2 Hz, 3 H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 200.5, 169.3, 167.4, 137.0, 136.0, 134.5, 129.3, 128.6, 127.1, 67.2, 63.0, 37.7, 31.9, 23.1, 14.1, 8.3.

*Diethyl 2-Acetamido-2-(3-butyrylbenzyl)malonate (**39**)*

Synthesized according to the general procedure with ketone **35** (8.0 g, 49.31 mmol), NBS (9.65 g, 54.24 mmol), AIBN (2.43 g, 14.79 mmol), and 123.28 mL CCl_4. The crude product was subjected to the next step without further purification.

Malonate synthesis was performed according to the general procedure using brominated **35** (5.73 g, 23.75 mmol), diethyl 2-acetamidomalonate (4.64 g, 21.37 mmol), K_2CO_3 (6.56 g, 47.5 mmol), KI (3.94 g, 23.75 mmol), and 148 mL MeCN. 37% yield, 2.9 g. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.82 (d, *J* = 7.8 Hz, 1 H), 7.63 (s, 1 H), 7.34 (t, *J* = 7.5 Hz, 1 H), 7.20 (d, *J* = 7.2 Hz, 1 H), 6.52 (s, 1 H), 4.28 (q, *J* = 7.2, 6.9 Hz, 4 H), 2.89 (t, *J* = 7.2 Hz, 2 H), 2.05 (s, 3 H), 1.75 (sx, *J* = 7.5, 7.2 Hz, 2 H), 1.31 (t, *J* = 7.2 Hz, 6 H), 0.99 (t, *J* = 7.5 Hz, 3 H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 200.1, 169.3, 167.4, 137.2, 136.0, 134.5, 129.4, 128.6, 127.1, 67.2, 63.0, 40.6, 37.7, 23.1, 17.8, 14.1, 14.0.

*Diethyl 2-Acetamido-2-(3-pentanoylbenzyl)malonate (**40**)*

Synthesized according to the general procedure with ketone **36** (9.5 g, 53.9 mmol), NBS (10.55 g, 59.29 mmol), AIBN (2.66 g, 16.17 mmol), and 134.75 mL CCl_4. The crude product was subjected to the next step without further purification.

Malonate synthesis was performed according to the general procedure using brominated **36** (6.26 g, 24.53 mmol), diethyl 2-acetamidomalonate (4.79 g, 22.08 mmol), K_2CO_3 (6.78 g, 49.07 mmol), KI (4.07 g, 24.53 mmol), and 153.3 mL MeCN. 51% yield as a yellow solid (4.4 g). ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.82 (d, *J* = 8.1 Hz, 1 H), 7.62 (s, 1 H), 7.36 (t, *J* = 7.8 Hz, 1 H), 6.52 (s, 1 H), 4.28 (q, *J* = 7.2 Hz, 4 H), 3.7 (s, 2 H), 2.92 (t, *J* = 7.5 Hz, 2 H), 2.05 (s, 3 H), 1.70 (p, *J* = 7.8, 7.5, 7.2 Hz, 2 H), 1.38 (p, *J* = 7.8, 7.5, 7.2 Hz, 2 H), 1.31 (t, *J* = 6.9, 7.2 Hz, 6 H), 0.95 (t, *J* = 7.2 Hz, 3 H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 200.2, 169.3, 167.4, 137.2, 135.9, 134.5, 129.3, 128.6, 127.1, 67.2, 62.9, 38.4, 37.7, 26.4, 23.1, 22.5, 14.0.

*Diethyl 2-Acetamido-2-(3-hexanoylbenzyl)malonate (**41**)*

Synthesized according to the general procedure with ketone **37** (10.0 g, 52.55 mmol), NBS (10.3 g, 57.81 mmol), AIBN (2.59 g, 15.77 mmol), and 131.38 mL CCl_4. The crude product was subjected to the next step without further purification.

Malonate synthesis was performed according to the general procedure using brominated **37** (5.73 g, 23.75 mmol), diethyl 2-acetamidomalonate (4.52 g, 20.79 mmol), K_2CO_3 (6.39 g, 46.21 mmol), KI (3.84 g, 23.11 mmol), and 144.44 mL MeCN. 37% yield as a yellow solid (3.1 g). ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.82 (d, *J* = 6.6 Hz, 1 H), 7.62 (s, 1 H), 7.36 (t, *J* = 7.5 Hz, 1 H), 7.20 (d, *J* = 7.5 Hz, 1 H), 6.52 (s, 1 H), 4.28 (q, *J* = 7.2 Hz, 4 H), 3.71 (s, 2 H), 2.91 (t, *J* = 7.5 Hz, 2 H), 2.05 (s, 3 H), 1.72 (m, 2 H), 1.34 (m, 4 H), 1.31 (m, 6 H), 0.91 (t, *J* = 6.6 Hz, 3 H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 200.2, 169.4, 167.4, 137.2, 135.9, 134.5, 129.3, 128.6, 127.1, 67.2, 62.9, 38.7, 37.7, 31.6, 24.0, 23.1, 22.6, 14.1.

Representative Procedure for Malonate Deprotection

*2-Amino-3-(3-propionylphenyl)propanoic Acid Hydrochloride (**12**)*

A suspension of **38** (1.0 g, 2.75 mmol) in 6 M HCl was refluxed overnight, until disappearance of protecting groups was verified via ^{1}H NMR. The resulting solution was concentrated *in vacuo* to yield **12** as a yellow solid (36% yield, 256 mg). ^{1}H NMR (300 MHz, D_2O) δ 7.93 (d, *J* = 6.9 Hz, 1 H), 7.87 (s, 1 H), 7.55 (m, 2 H), 4.16 (t, *J* = 7.2 Hz, 1 H), 3.30 (dq, *J* = 14.1, 5.7 Hz, 2 H), 3.09 (q, *J* = 7.2 Hz, 2 H), 1.14 (t, *J* = 6.9 Hz, 3 H). ^{13}C NMR (75 MHz, D_2O) δ 206.4, 172.4, 136.8, 135.1, 134.4, 129.3, 128.7, 127.5, 54.8, 35.7, 31.9, 7.5.

*2-Amino-3-(3-butyrylphenyl)propanoic Acid Hydrochloride (**13**)*

Synthesized according to the representative procedure using **39** (1.0 g, 2.65 mmol) in 9.14 mL 6 M HCl. 80% yield, 576 mg. ^{1}H NMR (300 MHz, D_2O) δ 7.95 (d, *J* = 7.2 Hz, 1 H), 7.85 (s, 1 H), 7.53 (m, 2 H), 4.28 (t, *J* = 7.2 Hz, 1 H), 3.31 (dq, *J* = 14.7, 5.7 Hz, 2 H), 3.02 (t, *J* = 7.2 Hz, 2 H), 1.66 (q, *J* = 7.2 Hz, 2 H), 0.92 (t, *J* = 7.2 Hz, 3 H). ^{13}C NMR (75 MHz, D_2O) δ 206.1, 171.5, 136.9, 134.8, 134.5, 129.4, 128.8, 127.8, 54.2, 40.4, 35.4, 17.6, 12.8.

*2-Amino-3-(3-pentanoylphenyl)propanoic Acid Hydrochloride (**14**)*

Synthesized according to the representative procedure using **40** (1.08 g, 2.75 mmol) in 9.48 mL 6 M HCl. 29% yield, 200 mg. ^{1}H NMR 7.96 (dt, *J* = 7.2 Hz, 1 H), 7.87 (s, 1 H), 7.58 (m, 2 H), 4.31 (t, *J* = 7.2 Hz, 1 H), 3.35 (dq, *J* = 14.7, 5.7 Hz, 2 H), 3.07 (t, *J* = 7.2 Hz, 2 H), 1.65 (p, *J* = 7.2 Hz, 2 H), 1.34 (sx, *J* = 7.5, 7.2 Hz, 2 H), 0.91 (t, *J* = 7.2 Hz, 3 H). ^{13}C NMR (75 MHz, D_2O) δ 206.2, 171.5, 136.9, 134.7, 134.5, 129.4, 128.8, 127.8, 54.2, 38.3, 35.4, 26.2, 21.6, 13.0.

*2-Amino-3-(3-hexanoylphenyl)propanoic Acid Hydrochloride (**15**)*

Synthesized according to the representative procedure using **41** (1.03 g, 2.55 mmol) in 8.79 mL 6 M HCl. 34% yield, 229 mg. ^{1}H NMR (300 MHz, D_2O) δ 7.95 (d, *J* = 7.2 Hz, 1 H), 7.87 (s, 1 H), 7.57 (m, 2 H), 4.22 (t, *J* = 7.2 Hz, 1 H), 3.33 (dq, *J* = 14.7, 7.8 Hz, 2 H), 3.08 (t, *J* = 7.2 Hz, 2 H), 1.68 (m, 2 H), 1.33 (m, 4 H), 0.86 (m, 3 H). ^{13}C NMR (75 MHz, D_2O) δ 205.2, 181.6, 139.1 136.3, 134.6, 128.8, 128.7, 126.3, 57.3, 40.9, 30.7, 23.8, 21.8, 13.3.

2.4.5. sfGFPS2TAG' Protein Sequence

MA**X**KGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFIC
TTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKD
DGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADK
QKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKD
PNEKRDHMVLLEFVTAAGITHGMDELYKGSHHHHHH

X denotes an amber stop codon for NAA incorporation in this study.

Compounds **5-15** were expressed using the sequence provided above. sfGFP expression of **1-4** used a sequence described previously.[115,116]

3. A GENETICALLY ENCODED ALDEHYDE FOR RAPID PROTEIN LABELING*

3.1. Introduction

Although the aldehyde is one of the most versatile chemical functionalities and participates in a variety of useful chemical reactions,[23,126-130] it is not a part of the 20 canonical amino acids. To harness the unique reactivity of aldehydes with hydrazine and hydroxylamine dyes for protein labeling, aldehyde tags have been introduced into proteins by oxidizing an N-terminal serine or via enzymatic modifications of preinstalled substrate peptides.[20,131-135] The N-terminal oxidation approach can only be applied *in vitro* and the enzymatic approach is generally confined for the aldehyde installation at the two protein termini. To genetically encode the keto functionality at any chosen site in a protein, several ketone-containing non-canonical amino acids (NCAAs) have been designed and incorporated into proteins in living cells using the amber suppression approach.[136-140] However, a method for the genetic incorporation of an aldehyde-containing NCAA has not been developed. We are interested in a genetically encoded, aldehyde-containing NCAA due to its superior reactivity with respect to ketone-containing NCAAs, and the observation that aniline can selectively enhance the reaction kinetics of the aldehyde–hydroxylamine/hydrazine condensation.[141-144]

---

*Tuley, A.; Lee, Y.-J.; Wu, B.; Wang, Z. U.; Liu, W. R. *Chemical Communications* **2014**, *50*, 7424. Reproduced by Permission of the Royal Society of Chemistry

3.2. Results and Discussion

In our previous studies, we revealed that a mutant pyrrolysyl-tRNA synthetase with mutations N346A/C348A (PylRS(N346A/C348A)) recognizes a number of phenylalanine derivatives including *meta*-acetylphenylalanine (**2** in Figure **3.1.A**) and, in coordination with $tRNA^{Pyl/CUA}$, directs their genetic incorporation at amber mutation sites in *Escherichia coli* and mammalian cells.[83,115,116,145,146] Given the size similarity between **2** and *meta*-formyl-phenylalanine (**1**), and the ability of this mutant enzyme to accept diverse *meta*-substituted phenylalanine substrates, we anticipated that this same

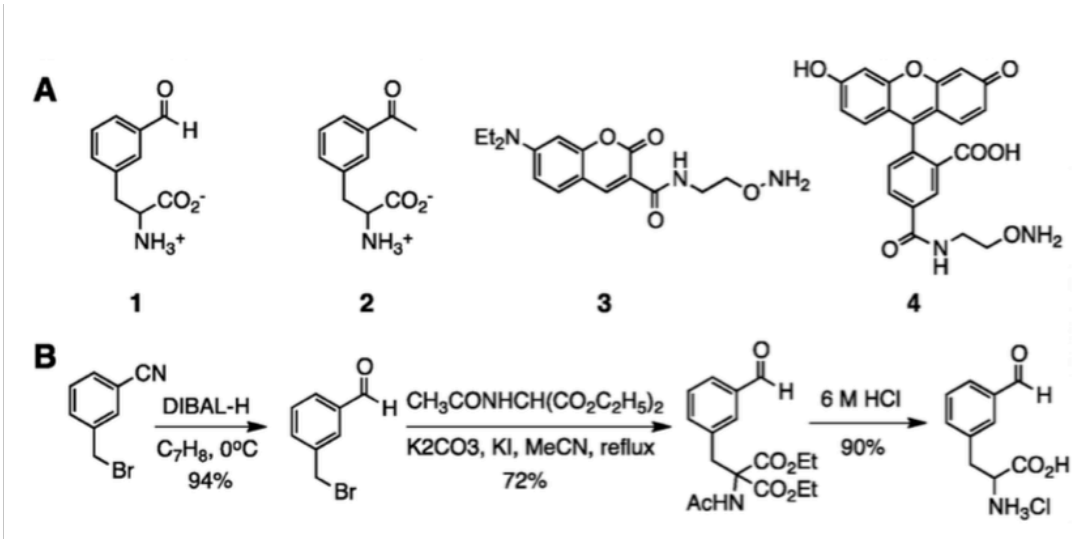


Figure 3.1. (A) Structures of **1–4**. (B) The synthetic route of **1**.

mutant enzyme would also accept **1** as its substrate. To test this prospect, **1** was readily prepared in a high yield by a three-step synthesis (Figure **3.1.B**). Its acceptance as a substrate of PylRS(N346A/C348A) was then examined. An *E. coli* BL21(DE3) cell that harbored two plasmids, pEVOL-pylT-PylRS(N346A/C348A) and pET-pylT-sfGFP2TAG, was employed for the investigation. Plasmid pEVOL-pylT-PylRS(N346A/C348A) contains genes coding both PylRS(N346A/C348A) and $tRNA^{Pyl/CUA}$. Plasmid pET-pylT-sfGFP2TAG carries a $tRNA^{Pyl/CUA}$ coding gene and a sequence-optimized superfolder green fluorescent protein (sfGFP) gene with an amber mutation at its S2 position. Growing cells in M9 minimal medium supplemented with 1% glycerol, 1 mM IPTG and 0.2% arabinose, but without **1**, afforded a minimal expression level of full-length sfGFP (<1 mg L^{-1}). Providing **1** at 2 mM into the medium promoted full-length sfGFP expression (Figure **3.2.A**). The expression level is comparable to that of para-propargyloxy-phenylalanine (ProY), and the electrospray ionization mass spectrometry analysis of the purified **1**-containing sfGFP displayed a molecular weight that agreed well with its theoretical value (Figure **3.2.B**). Moreover, compound **1** is not toxic to *E. coli* cells, as phenotypes and cell pellets are comparable to control experiments.

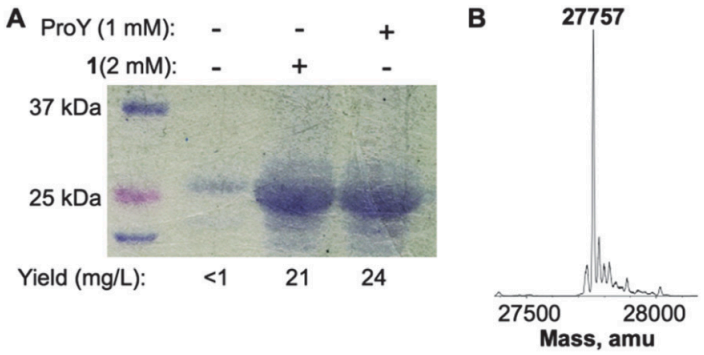


Figure 3.2. (A) The site-specific incorporation of **1** into sfGFP at its S2 position. Cells transformed with pEVOL-PylRS(N346A/C348A) and pET-pylT-sfGFP2TAG were grown under conditions with or without a NCAA supplement. (B) The deconvoluted ESI-MS spectrum of the purified **1**-containing sfGFP. Its theoretical value is 27 756 Da.

Dawson *et al.* previously showed that aniline could serve as an excellent catalyst for speeding the imine formation of aldehydes with hydroxylamines and hydrazines. To demonstrate this catalytically rapid reaction on proteins, sfGFP with **1** incorporated at its N149 position (sfGFP-**1**) was expressed and then applied to a Förster resonance energy transfer (FRET)-based kinetic investigation of its reaction with a coumarin dye **3**,[139,147] in the presence of 100 mM aniline and at pH 7 using a fluorescence spectrophotometer (Figure **3.1.A**). **3** has an excitation wavelength at 417 nm and emits at 476 nm. Its emitted light falls in the range of excitation lights of sfGFP that emits at 510 nm.[148] The reaction of sfGFP-**1** with **3** will place **3** in a close locality of the sfGFP fluorophore for FRET. All kinetic analyses were carried out under pseudo-first-order conditions in which **3** was at least in 5-fold excess in comparison to sfGFP-**1**. Data collected under these conditions were well fitted to a single exponential increase equation $F = F_1 - F_2 \, x \, e^{(-k' \, x \, t)}$, where F was the detected fluorescent signal at a given time, F_1 was the final

fluorescence, F_1-F_2 was the background fluorescent signal, and k' was the apparent pseudo first-order rate constant (Figure **3.3.A**). The determined k' values were plotted against the concentrations of **3** and fitted to the equation $k' = k\ [3] + C$, where k was a second-order rate constant for the reaction of sfGFP-**1** with **3** (inset of Figure **3.3.A**).

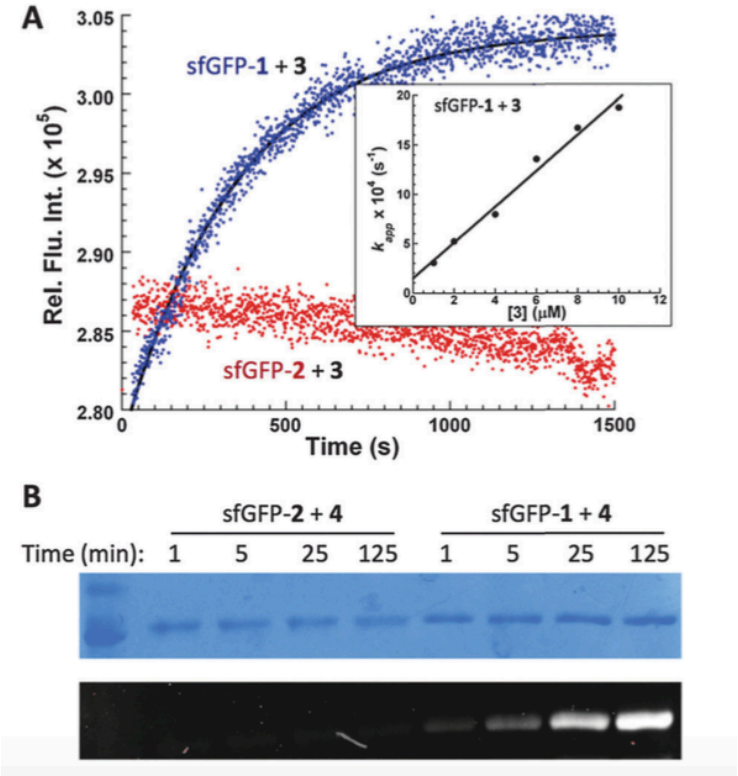


Figure 3.3. A) Fluorescent intensities as a function of time for reactions of 0.05 μM sfGFP-**1** (colored in blue) and 0.05 μM sfGFP-**2** (colored in red) with 10 μM **3** in the presence of 100 mM aniline and at pH 7. Presented in the inset is the linear dependence of the determined pseudo first-order rate constants for the reaction of 0.1 μM sfGFP-**1** with **3** on the concentrations of **3**. (B) Gel imaging analysis of labeling of sfGFP-**1** and sfGFP-**2** with **4** with different incubation times. For both sfGFP-**1** and sfGFP-**2**, labeling was carried out under conditions with 8 μM protein, 10 μM **4**, 100 mM aniline, and pH 7. Reactions were quenched with the addition of 2 mM benzaldehyde at indicated times and analyzed by denaturing SDS-PAGE. The top panel shows proteins with Coomassie blue staining and the bottom panel presents the fluorescent imaging of the same gel before Coomassie blue staining.

The calculated second-order rate constant was 182 ± 12 M^{-1} s^{-1}. This determined rate constant is comparable to what Dawson *et al.* reported for aniline-catalyzed bioconjugation, and is among one of the most rapid bioorthogonal reactions.[42,46,78,141-144,149-151] We also carried out the kinetic analysis in the absence of aniline. However, no obvious signal changes were observed when 0.1 μM sfGFP-**1** reacted with 4 μM **3** in a period of 2 h, indicating very slow kinetics of the sfGFP-**1** reaction with **3** in the absence of aniline and at pH 7. We also tried much higher concentrations of **3** to speed up the reaction rate. However, the background fluorescence from **3** under these conditions was too high to collect reliable data. As a comparison, a sfGFP variant with **2** incorporated at its N149 position (sfGFP-**2**) was also expressed and used for kinetic investigation of its reaction with **3** in the presence of 100 mM aniline and at pH 7. As shown in Fig. 3A, under the tested conditions of 0.05 μM sfGFP-**2** and 10 μM **3**, no obvious signal increase due to the sfGFP-**2** reaction with **3** was observed in a period of 30 min. The fluorescence decrease was caused by the bleaching effect of the excitation light.

The different labeling kinetics of sfGFP-**1** and sfGFP-**2** were also examined by fluorescent imaging of SDS-PAGE analyzed **4**-labelled products. Both sfGFP-**1** and sfGFP-**2** at 8 μM were incubated with 10 μM **4** in the presence of 100 mM aniline and at pH 7. Reactions were quenched at different times and analyzed by denaturing SDS-PAGE to remove the intrinsic fluorescence of sfGFP. Fluorescent imaging of two proteins at different labeling times clearly showed that sfGFP-**1** was efficiently labeled at 25 min but sfGFP-**2** was minimally labeled even at 2 h. Labeling could also be achieved in the absence of aniline, though reaction times were significantly longer (Fig.

3.6 in section 3.4.3). Once again, sfGFP-**1** showed superior labeling with respect to sfGFP-**2**. This result shows that, in systems where aniline may be toxic, labeling can still be achieved without it or at much lower concentrations.

With the demonstration of fast labeling kinetics of **1** with hydroxylamine dyes, we then proceeded to test the conditions for labeling proteins that bear site-specifically incorporated **1** in living cells. Plasmid pEVOL-pylT-PylRS(N346A/C348A) and a previously constructed plasmid, pETDuet-OmpXTAG, were used to transform *E. coli* BL21 cells. The latter plasmid contained a gene coding for an *E. coli* outer membrane protein OmpX,[78] with an AAAXAA (A denotes alanine and X denotes an amber mutation) insertion between two extracellular residues, 53 and 54. The transformed cells were grown in M9 minimal medium supplemented with 1% glycerol, 1 mM IPTG, 0.2% arabinose, and 2 mM **1** or **2** to express OmpX with **1** or **2** incorporated (OmpX-**1** or OmpX-**2**). Cells were then harvested and labeled with **4** in the presence of 100 mM aniline and at pH 7 for 1 h. After labeling, cells were washed with PBS buffer six times to remove the residual dye and then imaged by fluorescent microscopy. Cells grown in the absence of **1** or **2** were used as a control. As shown in Figure **3.4**, **4** specifically labeled cells expressing OmpX-**1** and OmpX-**2** were observed, but not for cells in the control experiment. Although cells expressing OmpX-**2** were fluorescently labeled, their intensities are weaker than those of cells expressing OmpX-**1**, indicating a slower labeling kinetics of OmpX-**2**. However, we have found that this concentration of aniline is toxic to *E. coli*, which underscores the need for developing less toxic catalysts for this reaction.

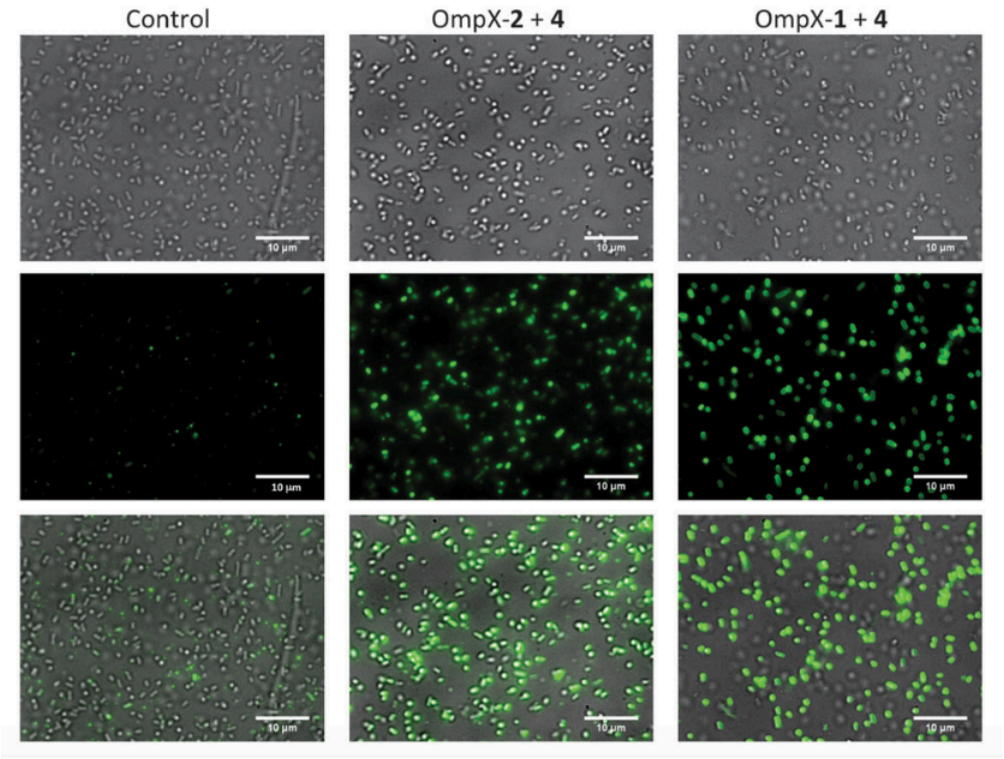


Figure 3.4. Selective labeling of *E. coli* cells expressing OmpX-**1** and OmpX-**2**. The labeling was carried out in the presence of 100 mM aniline and at pH 7 for 1 h. The top panel shows bright field imaging of *E. coli* cells, the middle panel shows green fluorescent imaging of same cells, and the bottom panel shows their composite images (scale bars are 10 μm).

3.3. Conclusion

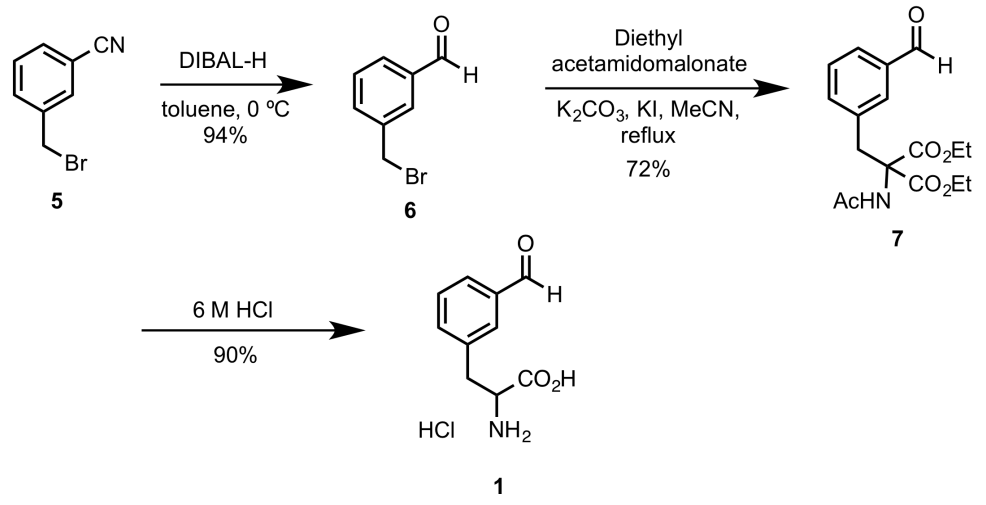
In summary, we reported the genetic incorporation of a readily synthesized aldehyde-containing NCAA and demonstrated its fast labeling kinetics with hydroxylamine dyes in the presence of the aniline catalyst. This rapid labeling approach was also successfully applied to label a membrane protein on the *E. coli* extracellular

surface. Although genetically encoded ketone-containing NCAAs were reported previously, labeling of proteins with these NCAAs typically suffers low labeling efficiencies, attributing to slow labeling kinetics of ketones. This work resolves this obstacle. Given that a large variety of hydroxylamine and hydrazine dyes are commercially available, as well as the existence of numerous aldehyde-based bioconjugation strategies,[15-17,152-161] we anticipate that this approach will be quickly adopted by others for studies such as protein folding/dynamics, protein–ligand interactions, high throughput drug discovery, etc.

3.4. Detailed Experimental Protocols

Organic synthesis was performed in oven-dried glassware under an argon atmosphere, and reagents were purchased and used without further purification. NMR spectra were obtained using Inova-300, Mercury-300, and Inova-500 instruments. Centrifugation was performed using a Sorvall Superspeed RC2-B automatic refrigerated centrifuge. Cell lysis was achieved with a VWR Scientific Branson Sonifier 450. Fluorometer data was obtained on an instrument purchased from Photon Technology International, Inc.; data from fluorometric studies were processed using Kaleidagraph. Gel imaging was achieved with a Bio-Rad ChemiDoc^{TM} XRS+ gel imager using ImgaeLab^{TM} software. Protein yield was determined using a Shimadzu UV-1800 spectrophotometer, using the sfGFP molar extinction coefficient of 83300 $M^{-1}cm^{-1}$.[148]

3.4.1. Organic Synthesis



*3-(Bromomethyl)benzaldehyde (**6**)*

Commercially available alpha-bromo *m*-toluinitrile **5** (5 g, 25.51 mmol) was dissolved in toluene and cooled to 0 ºC under an argon atmosphere. Diisobutylaluminum hydride (35.96 ml, 35.96 mmol, 1 M/Hexanes) was added dropwise to the solution, upon which the white opaque solution became clear and yellow. After 2 hrs, 66 ml of $CHCl_3$ was added to the solution, followed by 200 ml 10% HCl. After stirring for an additional hour, the organic layer was separated and washed with distilled water, then dried with Na_2SO_4 and concentrated to yield a clear oil, which crystallized overnight. The white crystals were filtered and washed with ice-cold hexanes, then allowed to dry to yield the desired compound in 94% yield (4.77 g). Spectroscopic data was in agreement with

literature values.[162] ^{1}H NMR (300 MHz, $CDCl_3$) δ 10.03 (s, 1H), 7.91 (s, 1H), 7.83 (d, *J* = 7.5 Hz, 1H), 7.68 (d, *J* = 7.5 Hz, 1H), 7.54 (t, *J* = 7.5 Hz, 1H), 4.55 (s, 2H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 191.6, 138.9, 136.8, 134.9, 129.7, 32.1.

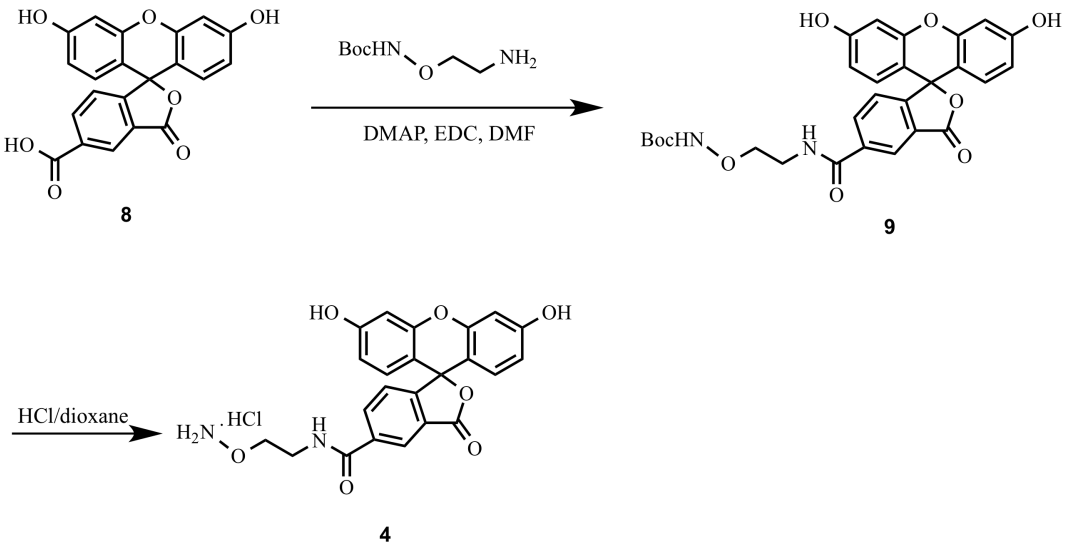
Diethyl 2-Acetamido-2-(3-formylbenzyl)malonate (7)

Alpha-bromo aldehyde **6** (4.52 g, 22.72 mmol), diethyl acetamidomalonate (4.44 g, 20.45 mmol), K_2CO_3 (6.28 g, 45.44 mmol), and KI (3.77 g, 22.72 mmol) were suspended in 142 ml of anhydrous MeCN and heated to reflux. After 15 hrs, the reaction was cooled to room temperature, filtered, and concentrated. The resulting crude residue was dissolved in boiling hexanes with minimal amounts of ethyl acetate; a yellow solid was obtained upon cooling, which was washed with ice-cold hexanes and dried *in vacuo* to yield **7** in 72% yield (5.4 g). ^{1}H NMR (300 MHz, $CDCl_3$) δ 9.97 (s, 1H), 7.76 (d, *J* = 7.5 Hz, 1H), 7.54 (s, 1H), 7.44 (t, *J* = 7.5 Hz, 1H), 7.29 (d, *J* = 7.5 Hz, 1H), 6.53 (s, 1H), 4.28 (q, *J* = 6.9, 7.2 Hz, 4H), 4.12 (q, *J* = 7.2 Hz, 1H), 3.75 (s, 2H), 2.05 (s, 3H), 1.31 (t, *J* = 7.2 Hz, 6H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 191.9, 169.3, 167.3, 136.5, 135.9, 130.3, 129.2, 129.0, 67.0, 62.9, 37.4, 23.0, 14.0.

*2-Amino-3-(3-formylphenyl)propanoic Acid Hydrochloride (**1**)*

m-Formyl malonate **7** from the previous reaction (5 g, 14.91 mmol) was dissolved in 51.4 ml of 6 M HCl, and the resulting suspension was refluxed for 33 hrs. The resulting solution was cooled to room temperature and the solvent was removed *in vacuo*, then the solid product was filtered and washed with diethyl ether to remove any organic impurities. The resulting off-white solid was lyophilized overnight to yield the title compound in 90% yield (3.08 g). ^{1}H NMR (300 MHz, D_2O) δ 9.79 (s, 1H), 7.78 (dt,

J = 1.8, 6.9 Hz, 1H), 7.72 (s, 1H), 7.46-7.55 (m, 2H), 4.19 (t, *J* = 6.9 Hz, 1H), 3.25 (dd, *J* = 6.3, 14.7 Hz, 3H). ^{13}C NMR (75 MHz, D_2O) δ 195.9, 171.2, 136.1, 136.0, 135.1, 130.3, 129.8, 129.7, 53.9, 35.2.



*tert-Butyl (2-(3',6'-dihydroxy-3-oxo-3H-spiro[isobenzofuran-1,9'-xanthene]-5-carboxamido)ethoxy)carbamate (**9**)*

To a solution of 5-carboxyfluorescein (0.10 g, 0.27 mmol), 4-(dimethylamino)pyridine (10 mg, 0.08 mmol) and *tert*-butyl (2-aminoethoxy)carbamate[117] (53 mg, 0.30 mmol) in anhydrous DMF (2 mL) was added *N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide hydrochloride (EDC hydrochloride, 78 mg, 0.41 mmol), and the mixture was stirred at room temperature overnight. The mixture

was diluted in ethyl acetate (50 mL), washed with sodium hydroxide (0.5 M, 10 mL), hydrochloric acid (0.5 *N*, 10 mL) and brine (10 mL), dried (Na_2SO_4), evaporated, and chromatographed (EtOAc/hexanes, 1:3 to 1:1) to give **9** (0.10 g, 70%) as a yellow oil. ^{1}H NMR (CD_3OD, 500 MHz) δ 8.95 (t, 1 H, *J* = 5.5 Hz), 8.49 (s, 1 H), 8.24 (dd, 1 H, *J* = 8.0, 1.5 Hz), 7.30 (d, 1 H, *J* = 8.0 Hz), 6.69 (d, 2 H, *J* = 2.0 Hz), 6.59 (d, 2 H, *J* = 9.0 Hz), 6.53 (dd, 2 H, *J* = 9.0, 2.5 Hz), 3.99 (t, 2 H, *J* = 5.2 Hz), 3.67 (t, 2 H, *J* = 5.0 Hz), 1.47 (s, 9 H); ^{13}C NMR (CD_3OD, 75 MHz) δ 170.7, 168.4, 161.5, 159.8, 156.9, 154.1, 137.8, 135.7, 130.3, 128.8, 125.8, 125.0, 113.8, 111.0, 103.8, 82.7, 75.9, 39.9, 28.7; HRMS (ESI) calcd for $C_{28}H_{27}N_2O_9$ ($[M+H]^+$) 535.1717, found 535.1234; calcd for $C_{28}H_{26}N_2O_9Na$ ($[M+Na]^+$) 557.1536, found 557.1028.

*N-(2-(Aminooxy)ethyl)-3',6'-dihydroxy-3-oxo-3H-spiro[isobenzofuran-1,9'-xanthene]-5-carboxamide Hydrochloride (**4**)*

To a solution of **9** (70 mg, 0.13 mmol) in 1,4-dioxane (1.0 mL) was added hydrogen chloride in dioxane (4.0 M, 0.3 mL, 1.2 mmol), and the mixture was stirred at room temperature for 4 h. The solvent was evaporated under high vacuum to afford **4** (62 mg, quant.) as a yellow oil. ^{1}H NMR (CD_3OD, 500 MHz) δ 8.78 (s, 1 H), 8.35 (s, 1 H), 7.57 (s, 1 H), 7.32 (bs, 2 H), 7.25 (bs, 2 H), 7.08 (bs, 2 H), 4.30 (bs, 2 H), 3.81 (bs, 2 H); HRMS (ESI) calcd for $C_{23}H_{19}N_2O_7$ ($[M+H]^+$) 435.1192, found 435.1292.

3.4.2. Superfolder Green Fluorescent Protein (sfGFP) Expression

sfGFPS2X Protein Sequence

MA**X**KGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKL
PVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTY
KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGI

KANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKR
DHMVLLEFVTAAGITHGMDELYKGSHHHHHH

X Denotes an amber stop codon where the non-canonical amino acid is incorporated.

sfGFPN149X Protein Sequence

MVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKL

PVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTY

KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSH**X**VYITADKQKNGI

KANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKR

DHMVLLEFVTAAGITHGMDELYKGSHHHHHH

X Denotes an amber stop codon where the non-canonical amino acid is incorporated.

sfGFPN149TAG Primers Sequences

sfGFPN149TAG-F: 5’ - TAG GTG TAT ATT ACC GCC GAT AAA CAG AAA AAT GG - 3’

sfGFPN149TAG-R: 5’ - ATG GCT GTT GAA ATT ATA TTC CAG TTT ATG - 3’

Construction of pBAD-sfGFPN149TAG

pBAD-sfGFP was a generous gift from Dr. Ryan Mehl at Franklin & Marshall College. sfGFP149TAG-F and sfGFP149TAG-R were used to perform the quick-change PCR on pBAD-sfGFP to afford pBAD-sfGFPN149TAG. This construct was cotransformed with pEVOL-PylT-PylRS-N346A/C348A into Top10 *E. coli* cells and used for sfGFP expression.

Genetic Incorporation of m-Formyl Phenylalanine Into Superfolder GFP

The procedure for expressing sfGFPS2TAG using BL21(DE3) *E. coli* has been reported previously.[115,116] For expressing sfGFPN149TAG, Top10 *E. coli* cells

containing pBAD-sfGFPN149TAG and pEVOL-PylT-PylRS-N346A/C348A were used to inoculate 5 ml of LB media supplemented with Cm (34 μg/ml) and Amp (100 μg/ml), then grown overnight, followed by transfer to 500 ml of LB media containing Cm (34 μg/ml) and Amp (100 μg/ml). When the cells had reached $O.D._{600}$ 1.3-1.6, the culture was spun down for 20 min at 4,000 rpm, then the resulting pellet was resuspended with water and spun down once more. The pellet from this centrifugation was resuspended in 15 ml ddH_2O, and 5ml of this suspension was added to a flask containing 45 ml minimal media (33.7 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 9.4 mM NH_4Cl, 1 mM $MgSO_4$, 0.3 mM $CaCl_2$, 1% glycerol) supplemented with 2 mM non-canonical amino acid (NCAA). Induction occurred via addition of 0.2% arabinose and 1 mM IPTG. After 12 hrs, the culture was centrifuged for 20 min at 4,000 rpm. The resulting pellet was resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8) and sonicated, then centrifuged for 1 hr at 4,000 rpm. The supernatant treated with imidazole to a final concentration of 10 mM, then incubated at 4 ºC for 1 h with Ni^{2+}-NTA resin. The resin was washed with 10 mM imidazole (3 x column volume) and 20 mM imidazole (3 x column volume), and then eluted with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 500 mM imidazole, pH 8). The protein was dialyzed three times with 20 mM Tris buffer, pH 8.1, and concentrated using Amicon Ultra-cel 10k centrifugal filter units. Protein purity was determined via SDS-PAGE analysis and subjected to ESI-MS.

Incorporation Using Autoinduction Media

Proteins were expressed in a manner similar to the protocol described above, the exception being the use of autoinduction media developed by Mehl and co-workers.[163]

The recipe used was identical to the referenced protocol except a different amino acid solution was utilized (400 μg/ml each of the following: glutamic acid sodium salt, aspartic acid, lysine hydrochloride, arginine hydrochloride, histidine hydrochloride monohydrate, alanine, proline, glycine, threonine, serine, glutamine, asparagine monohydrate, valine, leucine, isoleucine, tryptophan, and methionine). In other words, this 17 x amino acid solution differs from the reported recipe in that phenylalanine is removed to eliminate potential misincorporation. The cultures were allowed to grow overnight until a constant O.D. was obtained (1.82-1.83), upon which the cells were spun down, resuspended in lysis buffer, and sonicated. After centrifugation for 1 hr at 4,000 rpm, the lysate was purified and analyzed as described above, yielding full-length sfGFP.

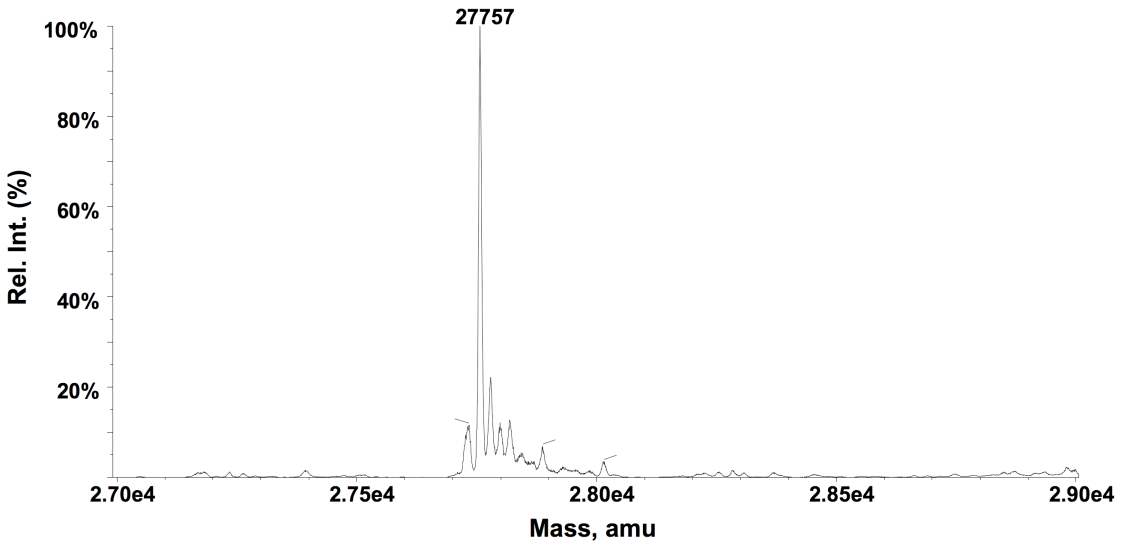


Figure 3.5. ESI-MS data for sfGFPS2X-**1**, M9 minimal media. Predicted mass: 27756. Observed mass: 27757.

3.4.3. In Vitro Aniline-Catalyzed Labeling of sfGFP

Determination of Second-Order Rate Constant

A stock solution of sfGFP was added to a solution of 100 mM aniline and 100 mM sodium phosphate, pH 7, to a total volume of 2 ml. To this solution was added coumarin-based dye **3**, and the reaction rate was monitored with an excitation wavelength of 417nm and emission of 510 nm for 2 hours. This reaction was repeated with various concentrations of hydroxylamine dye while keeping the sfGFP concentration constant, and each resulting scatterplot was fitted to an exponential trendline. The observed rate constants obtained from these runs were plotted, and the second-order rate constant was derived as described in the manuscript.

Time-Based Fluorescence Assay

An aliquot of sfGFP-**1** was diluted with a 2X stock of labeling buffer (200 mM aniline, 200 mM sodium phosphate, pH 7), with a final concentration of 8 μM sfGFP, 100 mM aniline, and 100 mM sodium phosphate. To this was added 10 μM of fluorescein-based hydroxylamine dye. The reaction was quenched with 2 mM benzaldehyde and immediately treated with SDS-PAGE loading buffer, and the resulting samples were subjected to 12% SDS-PAGE analysis. Quenching occurred after 1, 5, 25, and 125 minutes to give the appropriate samples for the assay. The above procedure was also used for sfGFP-**2**.

Fluorescence Assay Without Aniline Catalyst

To a solution of sfGFP-**1** in PBS buffer (8 μl, 3.6 μM) was added dye **4** (1.8 μl, 20 μM) to a final concentration of 2.9 μM sfGFP and 3.6 μM dye **4**. After the designated

amount of time had elapsed (1 h, 2 h, 7 h, and 18 h), the solution was precipitated using trichloroacetic acid, then the resulting pellet was washed twice with ice-cold acetone. After allowing the pellet to dry, it was resuspended in 20 μl H_2O and prepared for SDS-PAGE analysis using conventional methods. An identical procedure was performed for sfGFP-**2**, resulting in a total of eight reactions.

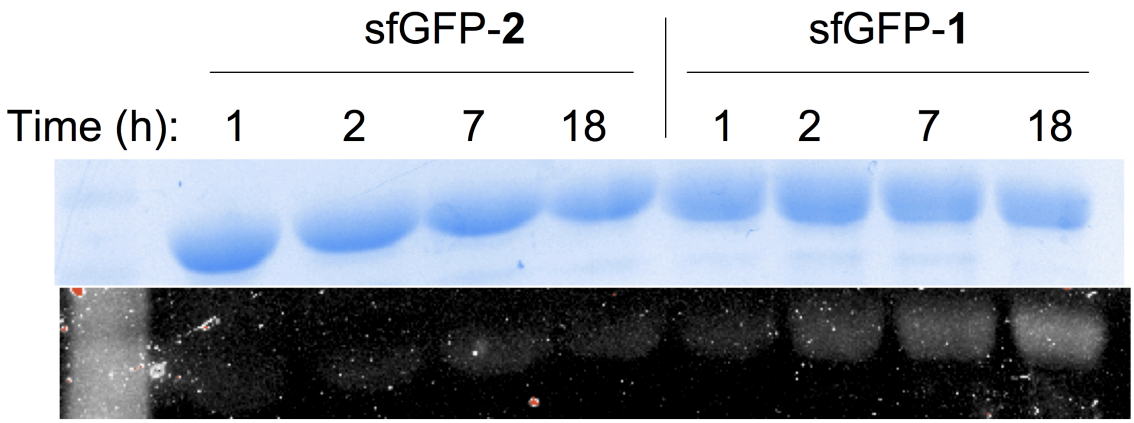


Figure 3.6. Bioconjugation of aldehyde in the absence of aniline catalyst at neutral pH. No appreciable labeling was observed with ketone-containing sfGFP-**2**.

3.4.4. In Vivo Labeling of OmpX

OmpX Expression

Expression of the membrane protein OmpX was performed as described previously.[78] Briefly, BL21(DE3) cells cotransformed with pEVOL-pylT-N346A/C348A and pETDuet-OmpXTAG were grown in 500 mL LB media until the O.D. reached 1.0-1.3, then the cells were pelleted and washed three times with PBS buffer. The cells were resuspended and transferred to 45 mL of minimal media supplemented with 2 mM NCAA, 1 mM IPTG, and 0.2% arabinose. Protein was expressed at 23 ºC for 10 h, and then the cells were pelleted and labeled as described below.

In Vivo Labeling

A 50 μL suspension of BL21(DE3) *E. coli* cells from procedure **4.1** was spun down for 5 min at 14,000 r.p.m., then resuspended in 100 μL labeling solution (100 mM aniline, 100 mM sodium phosphate, pH 7). Labeling was achieved by supplementing the suspension with 2 mM Dye **4**, and the reaction was left at 23 ºC for 1 h. When complete, the cells were spun down for 5 min at 14,000 r.p.m., then resuspended in PBS buffer and centrifuged again. The previous step was repeated for a total of six PBS washes, and the pellet was resuspended once more in PBS buffer and visualized via confocal microscopy.

4. DEVELOPMENT OF CHEMICALLY MODIFIED PHAGE DISPLAY LIBRARIES

4.1. Introduction

Directed evolution is a powerful approach to discover useful mutants of a protein or peptide of interest, whereby extraordinarily large numbers of individuals can be screened simultaneously and selected against a property of interest.[164] One of the requirements of a directed evolution study is that genotype and phenotype must be linked together. In other words, a given phenotype must be isolable and amplified using the specific genetic information that codes for such a phenotype. A prominent subset of directed evolution approaches include the "display" technologies, whereby a peptide or protein of interest is physically linked to the genetic information that codes for said peptide or protein. These various display systems have been developed for the *in vitro* selection of proteins, and each technique has its own method of linking phenotype to genotype. For example, techniques such as ribosome display and mRNA display rely on purified translational machinery to produce proteins; protein and mRNA are physically linked in an artificial manner. Conversely, other methodologies utilize entire microorganisms to link genotype and phenotype; these latter approaches include bacterial display, yeast display, and phage display.

Of the myriad techniques available for *in vitro* protein selection, phage display is among the most common approaches because its simplicity facilitates use by non-specialists.[165] George P. Smith initially reported phage display in 1985,[166] wherein foreign peptides were expressed on the outer protein coat of filamentous phage. Since that initial report, phage display libraries are routinely utilized for evolving a variety of

antibodies, enzymes, and peptide-based inhibitors. Although peptide-based inhibitors are a useful class of compounds, the utilization of display technologies for their selection means these inhibitors are delegated to the twenty canonical amino acids, leaving them inherently limited from a chemical standpoint. Further, peptide-based inhibitors lack the stability and bioavailability that smaller molecules possess, which limits their medicinal applications. However, cyclization of peptides could circumvent these issues, and this task could be achieved via bioorthogonal reagents. The first attempts to expand chemical diversity in phage utilized bioconjugation techniques, notably with nucleophilic cysteine residues. However, installation of bioorthogonal reagents would negate chemoselectivity concerns when appending functionality to phage. Such bioorthogonal labeling would allow for functionalization with pharmacophores that may be inaccessible through straightforward genetic code expansion, such as β-lactones, epoxides, and aziridines.

Schultz was the first to report successful genetic code expansion in phage systems using azidophenylalanine (AzF),[167] verifying that non-canonical functionality could be directly incorporated into phage in a site-selective manner. Although this pioneering work demonstrates that a system such as phage display can tolerate expansion of the genetic code, the amber stop codon is delegated to a specific location at the N-terminal peptide. In order to take full advantage of an expanded genetic code in directed evolution experiments, it would be necessary to develop a means by which one could randomize the location of the amber stop codon. Development of a TAG-enriched phage display library is challenging, because phage harboring the amber codon will be reproduced at a lower rate than phage that are either wild type or lack in-frame amber

codons. Phage that possess in-frame amber codons will grow and propagate at a slower rate because of inherent inefficiencies of genetic code expansion, and translation with canonical incorporation machinery will be more facile than non-canonical amino acid incorporation. Over time, the number of phage containing in-frame amber codons will diminish and become a smaller representation of the overall population. Consequently, only a small percentage of phage with the desired modification would be present during selection, resulting in a low probability that selected mutants would have non-canonical amino acids.

4.2. Progress Towards Developing a Non-Canonical Phage Display System

In order to address the above issues, our lab developed a method by which phage display libraries could be enriched in amber stop codons (Dr. Catrina Reed, unpublished results, Figure **4.1**). The initial, naïve library was superinfected with CM13 phage, and pADLg3 plasmids capable of producing full-length P3 were immune to the superinfection. Phage susceptible to the superinfection (i.e. TAG-encoding or otherwise non-functional mutants) acquired the CM13 phagemid, and these individuals were selected via growth in the presence of Kan and Amp. The DNA from these surviving cells was isolated and used to transform DH5α cells harboring M13K07g3TAA, a helper phagemid with an in-frame ochre codon (TAA) to prevent competitive expression of wild type P3 protein. DH5α is an amber-suppressing strain, meaning it possesses translation machinery capable of reading through in-frame amber codons. Consequently, phage expression using this system led to library enrichment in amber codons, because individuals in the library that possessed other types of deleterious mutations were not

able to produce phage. Sequencing of the resulting library revealed that ten out of ten selected clones possessed an in-frame amber codon.

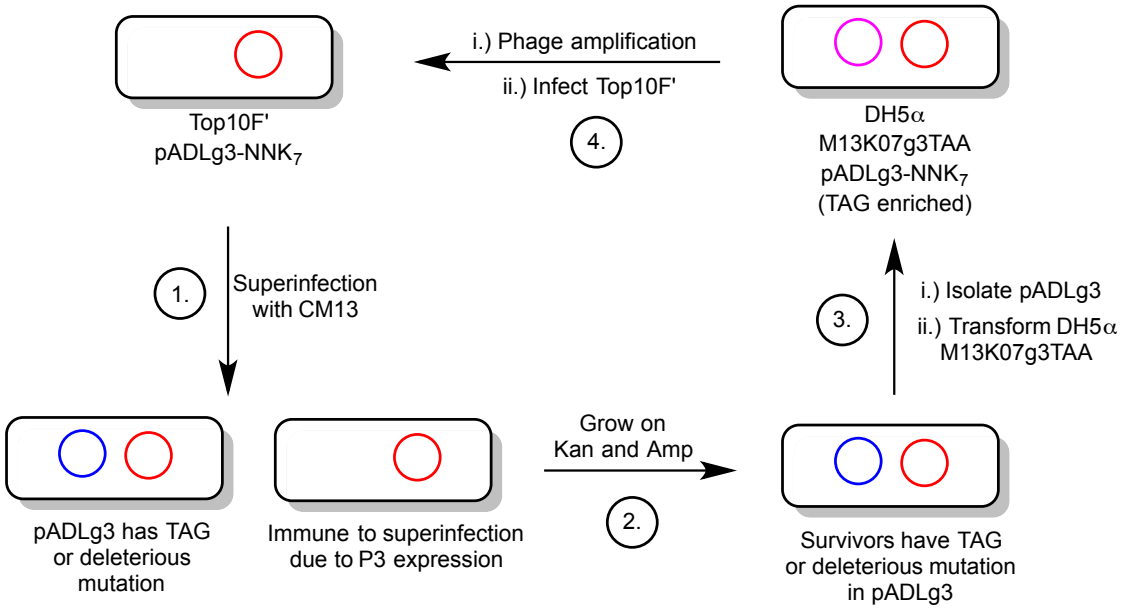


Figure 4.1. Strategy to acquire TAG-enriched phage libraries (Dr. Catrina Reed). Red plasmid: pADLg3-NNK_7 library. Blue plasmid: CM13 phagemid. Purple plasmid: M13K07g3TAA helper phagemid.

The resulting TAG-enriched library, pADLg3-$(NNK)_7$, was used to transform Top10 *E. coli* cells harboring two additional plasmids: pEVOL-(ClodF)-AzFRS and M13K07g3TAA. The pEVOL plasmid possesses a gene encoding azidophenylalanyl-tRNA synthetase (AzFRS) and has a unique ClodF origin of replication to distinguish this plasmid from the competing origins in the other plasmids. M13K07g3TAA is a

helper phage, providing all components necessary for phage production, with the exception of the P3 protein (the g3 gene that encodes P3 has an in-frame TAA codon to prevent translation in this system); consequently, amplified phage must acquire P3 from the pADL plasmid, resulting in all viable phage to be within inclusion of the randomized $(NNK)_7$ library.

Using this library, it was envisioned that phage could be functionalized with useful electrophilic moieties prior to selection. In order to demonstrate this technique, a β-lactone probe was synthesized; this lactone is derived from threonine, and was subsequently acylated by carboxylic acids possessing terminal alkynes.[168,169] Threonine-based lactones were chosen because they possess the ring strain (and reactivity) inherent to β-lactones, but are stable enough to tolerate labeling and panning conditions. Furthermore, this alkyne-functionalized β-lactone reacted sufficiently with azido-coumarin, demonstrating its utility as a bioconjugation reagent (Figure **4.2**).

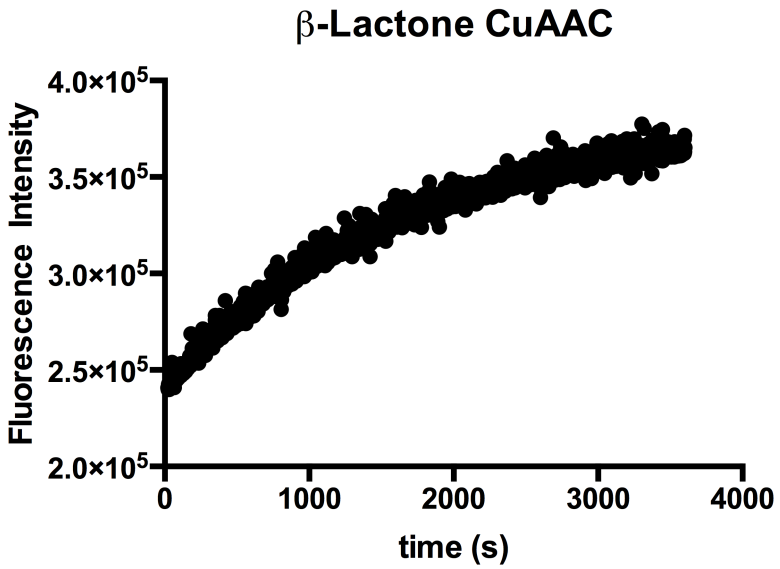


Figure 4.2. Copper-catalyzed azide-alkyne cycloaddition with β-lactone and azido-coumarin. Conditions: 2 mM β-lactone probe, 50 μM $CuSO_4$, 300 μM BTTP, 2.5 mM sodium ascorbate, 5% DMSO, 100 mM potassium phosphate.

With phage and probe in hand, it was then necessary to survey conditions that would permit bioconjugation while simultaneously maintaining the infectivity of phage. Since infectivity is crucial for amplification, potential binders may be lost simply because they could not be amplified after elution. Individually, each reagent had no impact on infectivity with respect to control experiments, but samples containing all reagents saw orders of magnitude less infectivity. However, through the use of copper-chelating BTTAA, infectivity was maintained when treated under click chemistry conditions for one hour in PBS buffer. As a proof of principle, efforts are underway to select β-lactone based binders of tobacco etch virus (TEV) protease utilizing this approach (Figure **4.3**); one round of selection has been attempted using a surface

panning procedure, but no phages were detected in the elution by measuring the number of colony forming units (cfu). More optimization is needed with regards to the bioconjugation and panning procedures.

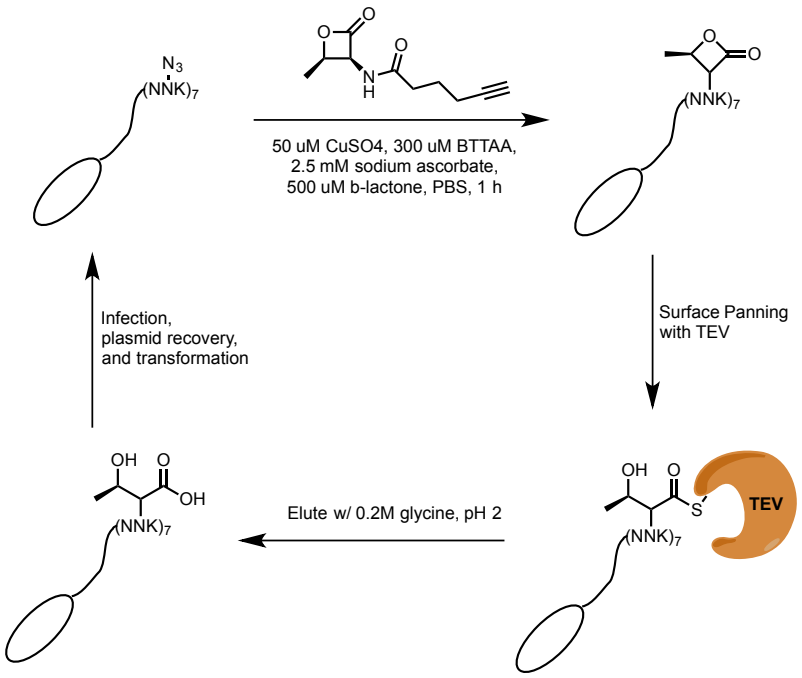


Figure 4.3. Strategy for functionalization and selection of phage possessing β-lactones.

4.3. Development of a Thiazolidine-Based Phage Display Library

The lack of chemical diversity is not the only limitation to phage display libraries. All inhibitors discovered via phage display would be peptide-based, and this class of compounds suffers from proteolytic susceptibility and poor cell permeability,

both of which lower bioavailability. Furthermore, selectivity is often an issue since peptides can achieve a number of conformations that allow them to interact with multiple targets. Peptide cyclization could potentially circumvent such issues,[170-172] and methods to cyclize phage display libraries have been pursued in recent years.[173,174]

Among the approaches currently being studied, we hypothesized that aldehyde-based linkages could be a useful means of constructing cyclic peptide libraries. In the presence of an N-terminal cysteine, aldehydes could spontaneously form thiazolidines, a stable heterocycle at neutral pH.[175,176] In order to determine if this approach would be viable, a superfolder green fluorescent protein (sfGFP) fusion protein was constructed with a leader peptide possessing an N-terminal tobacco etch virus (TEV) recognition sequence, followed by cysteine, five consecutive alanines, and a genetically encoded aldehyde (3-formyl-phenylalanine, designated as "RCHO"). The resulting TEV-Cys$(Ala)_5$-RCHO-sfGFP protein was expressed, purified, and tested for cyclization (Figure **4.4**). Upon cleavage of the TEV recognition site, the resulting N-terminal cysteine residue should be capable of condensing with 3-formyl phenylalanine to form the thiazolidine. In order to test this cyclization, the Cys$(Ala)_5$-TAG-sfGFP protein was treated with a cyanobenzonitrile dye selective for N-terminal cysteines.[177] If the N-terminal cysteine does not react with aldehyde, it should be available for condensation with the cyanobenzonitrile dye. However, no such labeling was observed, and several controls were simultaneously assayed, including an attempt to label without TEV cleavage, and repetition of both experiments on the similar TEV-Cys$(Ala)_5$Cys-sfGFP protein. Only one protein, Cys$(Ala)_5$Cys-sfGFP, was successfully labeled.

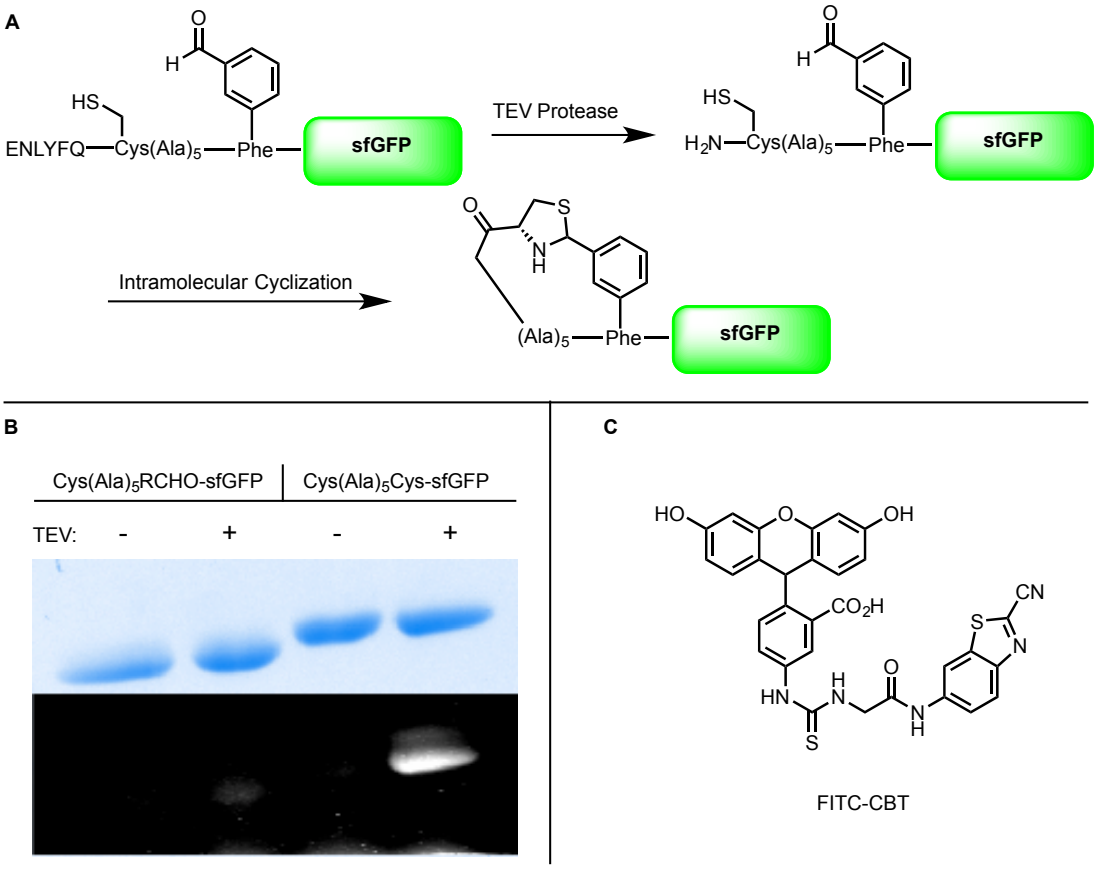


Figure 4.4. (A) General scheme for TEV cleavage and intramolecular cyclization. (B) SDS-PAGE analysis of proteins stained with Coomassie brilliant blue (top) and fluorescent imaging (bottom) demonstrating that only $Cys(Ala)_5Cys$-sfGFP is labeled after TEV cleavage. Conditions: 8 μM GFP, 0.8 μM TEV protease, 50 mM sodium phosphate, 200 mM NaCl. **C.** Structure of FITC-CBT.

The results discussed above were sufficient preliminary data to move forward with construction of a thiazolidine-based phage display library. Commercially available pADLg3 (Antibody Design Labs) containing full length P3 was subjected to PCR in order to install the necessary randomized sequence on the N-terminus of P3. Top10 *E.*

coli cells were transformed with the resulting pADLg3-Cys-$(NNK)_7$-TAG plasmid, and the number of transformants was determined by plating 10-fold dilutions of the transformed cells onto agar plates containing ampicillin, and quantify the number of colony forming units (theoretical library size: 1.28×10^9, actual library size: 4.26×10^9). Currently, the constructed library has not been assessed for biases. However, given that the actual library size is greater than three times the theoretical size, it is likely that adequate representation of each possible individual has been obtained.

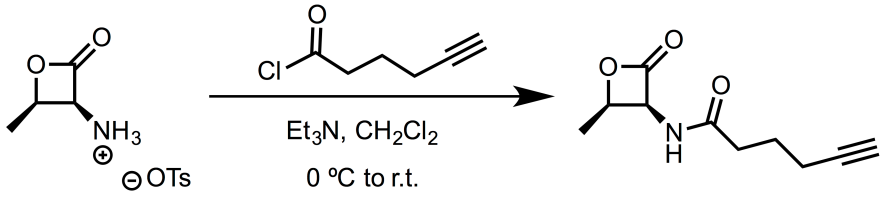
4.4. Conclusion

The ability to obtain phage libraries enriched in amber codons has paved the way for developing a number of methodologies to expand the chemical diversity of phage. With TAG-enriched libraries in hand, one can perform directed evolution experiments wherein the non-canonical amino acid occurs in all possible locations within the peptide sequence, allowing for greater surveillance of the fitness landscape. Further, peptide cyclization via thiazolidine formation appears to be a plausible approach to obtain cyclic peptides, and work is underway to translate these preliminary protein trials to phage display systems. It is anticipated that the preliminary work reported herein will benefit those who wish to pursue methodologies involving chemically modified phage, a promising approach to discovering novel therapeutic peptides.

4.5. Detailed Experimental Protocols

Organic synthesis was performed in oven-dried glassware under an argon atmosphere, and reagents were purchased and used without further purification. NMR spectra were obtained using Inova-300 and Mercury-300 instruments. All media was autoclaved and/or sterile filtered prior to use. Cloning and library construction utilized reagents and protocols obtained from New England Biolabs. For experiments involving phage, all pipettes and bench tops were washed thoroughly with 70% ethanol after use. Centrifugation was performed using a Sorvall Superspeed RC2-B automatic refrigerated centrifuge. Cell lysis was achieved with a VWR Scientific Branson Sonifier 450. Fluorometer data was obtained on an instrument purchased from Photon Technology International, Inc.; data from fluorometric studies were processed using Kaleidagraph. Gel imaging was achieved with a Bio-Rad ChemiDoc^{TM} XRS+ gel imager using ImgaeLab^{TM} software. Protein yield was determined using a Shimadzu UV-1800 spectrophotometer, using the sfGFP molar extinction coefficient of 83300 $M^{-1}cm^{-1}$.

4.5.1. Organic Synthesis



N-((2R,3S)-2-Methyl-4-oxooxetan-3-yl)hex-5-ynamide

Threonine-based β-lactone was synthesized according to an established literature procedure.[178,179] To a suspension of this β-lactone (50 mg, 0.183 mmol) in dichloromethane (1.02 mL) was added triethylamine (0.1 mL, 0.732 mmol) dropwise, at which point the suspension became a clear solution. Hexynoyl chloride (35.78 mg, 0.274 mmol) was added dropwise as a 1 M solution in dichloromethane at 0 ºC, and allowed to warm to room temperature overnight. After 16 h, the crude reaction mixture was concentrated *in vacuo* and immediately purified via silica gel chromatography (75% ethyl acetate/hexanes) to yield a white solid in 71% yield (25.2 mg). ^{1}H NMR (300 MHz, $CDCl_3$) δ 6.46 (s, 1 H), 5.62 (m, 1 H), 4.90 (p, 1 H, J = 6.3 Hz), 2.43 (t, 2 H, J = 7.5 Hz), 2.26 (m, 2 H), 2.00 (s, 1 H), 1.89 (p, 2 H, J = 7.2 Hz), 1.43 (d, 3 H, J = 6.3 Hz). ^{13}C NMR (75 MHz, $CDCl_3$) δ 172.4, 169.4, 83.1, 75.0, 69.7, 58.9, 34.3, 23.8, 17.8, 15.1.

4.5.2. pADLg3-(NNK)$_7$ (TAG Enriched)

pADLg3-(NNK)$_7$ (TAG Enriched) Transformation

To an ice-cold 0.6 mL tube was added 100 ng of pADLg3-Cys(NNK)$_7$ (100 ng/μL), and a suspension of electrocompetent Top10 cells (99 μL) harboring M13K07g3TAA and pEVOL-ClodF-AzFRS was slowly added by gently swirling the pipette while dispensing (100 μL total). From here, the suspension was transferred to a clean electroporator cuvette with a 2 mm gap. This procedure was repeated fifteen times. After each electroporation, cells were recovered by quickly adding 900 μL of pre-warmed (37 ºC) 2YT, and this 1 mL culture was recovered for 1 h at 37 ºC. Then, the individual recovered cultures were pooled and grown in 500 mL of 2YT supplemented

with Kan (50 μg/mL), Amp (100 μg/mL), and Cm (34 μg/mL) overnight. Finally, cells were harvested by centrifugation for 20 min at 4,000 rpm, followed by resuspension in 50 mL 20% glycerol/2YT supplemented with Kan (50 μg/mL), Amp (100 μL/mL), and Cm (34 μg/mL). Aliquots were stored at -80 ºC to be used later for phage expression.

TAG-Enriched Phage Expression

A 40 μL aliquot of *E. coli* Top10 cells containing helper phage M13K07g3TAA, synthetase plasmid pEVOL-(ClodF)-AzFRS, and pADLg3-Cys$(NNK)_7$TAG library was used to inoculate 200 mL of 2YT media supplemented with Kan (50 μg/mL), Amp (100 μg/mL), and Cm (34 μg/mL). The initial optical density at 600 nm ($O.D._{600}$) was 0.04, and this culture was grown at 37 ºC until the $O.D._{600}$ equaled 0.56, then the culture was divided into two 100 mL samples (one as a control, another supplemented with AzF). P3 and AzFRS expression was induced at 30 ºC for 12 h, with the addition of 0.2% arabinose, 0.5 mM IPTG, and 2 mM AzF. After 12 h, cells were centrifuged for 20 min at 4,000 rpm, and the supernatant was recovered. The phage were precipitated by addition of 20 mL of PEG precipitation solution (20% polyethylene glycol-8000, 2.5 M NaCl) and left overnight at 4 ºC. The next day, precipitated phage were recovered by centrifugation for 25 min at 11,000 x g, and the phage were resuspended in 5 mL phosphate buffered saline (PBS) by gently rocking at 4 ºC for 3 h. Residual *E. coli* were removed via centrifugation for 20 min at 4,000 rpm, the supernatant was transferred to a clean tube, and 2 mL PEG precipitation solution was added. After overnight precipitation at 4 ºC, phage were once again recovered by centrifugation for 25 min at 11,000 x g, then resuspended in 1 mL PBS by gently pipetting. Residual *E. coli* were

removed via centrifugation for 20 min at 4,000 rpm, and the supernatant was transferred to a clean 1.5 mL Eppendorf tube. After heating to 65 ºC for 15 min, the purified phage were stored at 4 ºC until needed. The amount of phage present was determined by titering: phage were diluted 1,000-fold, used to infect Top10 F' cells ($O.D._{600}$ = 0.4-0.6) for 45 min at 37 ºC, then these infected cells were diluted several orders of magnitude and grown on an Amp plate (100 µg/mL).

Determining Reagent Effects on Infectivity

100 µL of a CM13 phage suspension were diluted in 900 µL PBS containing 50 µM $CuSO_4$, 50 µM sodium ascorbate, 50 µM β-lactone, and 300 µM BTTAA. After incubation for 1 h, 100 µL of this solution was used to infect Top10 F' cells ($O.D._{600}$ = 0.4-0.6) for 45 min, followed by serial dilution up to 10^{-8}. 10 µL of each dilution was spotted onto an LB agar plate supplemented with Kan (50 µg/mL) and grown overnight at 37 ºC. As a control, a sample of CM13 phage was also subjected to 1 h PBS only (no click reagents) and titered in the same manner.

4.5.3. Thiazolidine-based Peptide Cyclization and Library Construction

sfGFP-Cys(Ala)$_5$-TAG Primers

CA5TAG Fwd:

5' – CGG CCG CGG CCT AGG TTA GCA AAG GTG AAG AAC TG – 3'

CA5TAG Rev:

5' – CGG CGC ACT GAA AAT ACA GGT TTT CCA TGG TTA ATT CCT CCT G – 3'

sfGFP-Cys(Ala)$_5$-TAG Sequence

DNA Sequence

atggaaaacctgtattttcagtgcgccgcggccgcggcctaggttagcaaaggtgaagaactgtttaccggcgttgtgccgatt

ctggtggaactggatggtgatgtgaatggccataaatttagcgttcgtggcgaaggcgaaggtgatgcgaccaacggtaaact

gaccctgaaatttatttgcaccaccggtaaactgccggttccgtggccgaccctggtgaccaccctgacctatggcgttcagtgc

tttagccgctatccggatcatatgaaacgccatgatttctttaaaagcgcgatgccggaaggctatgtgcaggaacgtaccattag

cttcaaagatgatggcacctataaaacccgtgcggaagttaaatttgaaggcgataccctggtgaaccgcattgaactgaaagg

tattgattttaaagaagatggcaacattctgggtcataaactggaatataatttcaacagccataatgtgtatattaccgccgataaa

cagaaaaatggcatcaaagcgaactttaaaatccgtcacaacgtggaagatggtagcgtgcagctggcggatcattatcagca

gaataccccgattggtgatggcccggtgctgctgccggataatcattatctgagcacccagagcgttctgagcaaagatccgaa

tgaaaaacgtgatcatatggtgctgctggaatttgttaccgccgcgggcattacccacggtatggatgaactgtataaaggcagc

caccatcatcatcaccattaa

Note: Amber stop codon highlighted in red.

Peptide Sequence

M E N L Y F Q C A A A A A A **X** V S K G E E L F T G V V P I L V E L D G D V N G

H K F S V R G E G E G D A T N G K L T L K F I C T T G K L P V P W P T L V T T

L T Y G V Q C F S R Y P D H M K R H D F F K S A M P E G Y V Q E R T I S F K D

D G T Y K T R A E V K F E G D T L V N R I E L K G I D F K E D G N I L G H K L E

Y N F N S H N V Y I T A D K Q K N G I K A N F K I R H N V E D G S V Q L A D H

Y Q Q N T P I G D G P V L L P D N H Y L S T Q S V L S K D P N E K R D H M V L

L E F V T A A G I T H G M D E L Y K G S H H H H H H

X denotes the location of a non-canonical amino acid.

sfGFP-Cys(Ala)$_5$-TAG Cloning Procedure

The above primers were used to perform a 50 µL scale PCR on pBAD-sfGFP-Cys-Ala$_5$-Cys with Phusion DNA polymerase. The amplified product was subjected to DpnI digestion for 2 h at 37 ºC, followed by ligation with T4 DNA ligase for 4 h at 23 ºC. The ligated product was used to transform chemically competent Top10 cells, followed by growth on an Amp plate (100 µg/mL), and several colonies were selected for sequencing. The resulting plasmid was used with pEVOL-PylRS-N346A/C348A to co-transform chemically competent Top10 cells. After growth on an Amp/Cm plate, one colony was chosen to prepare cell stocks.

Cys(Ala)$_5$-RCHO-sfGFP Expression Procedure

A 5 mL culture of *E. coli* Top10 cells containing pBAD-sfGFP-Cys-Ala$_5$-TAG and pEVOL-PylRS-N346A/C348A were grown in Amp (100 µg/mL) and Cm (34 µg/mL) overnight, then used to inoculate 500 mL autoinduction media (the recipe used was identical to the referenced protocol except the amino acid solution utilized 400 µg/ml each of the following: glutamic acid sodium salt, aspartic acid, lysine hydrochloride, arginine hydrochloride, histidine hydrochloride monohydrate, alanine, proline, glycine, threonine, serine, glutamine, asparagine monohydrate, valine, leucine, isoleucine, tryptophan, and methionine) with Amp (100 µg/mL) and Cm (34 µg/mL), supplemented with 2 mM 3-formyl phenylalanine. After the O.D.$_{600}$ became constant (~5, 24 h) the cells were centrifuged for 20 min at 4,000 rpm, followed by sonication and subsequent centrifugation at 10,000 rpm for 1 h. The resulting sfGFP fusion protein was purified from the soluble fraction via nickel affinity chromatography (3x CV 10 mM

imidazole, 3x CV 20 mM imidazole), dialyzed against 10 mM ammonium bicarbonate (pH 8) and concentrated via speed vac to yield full-length protein (16 mg/L).

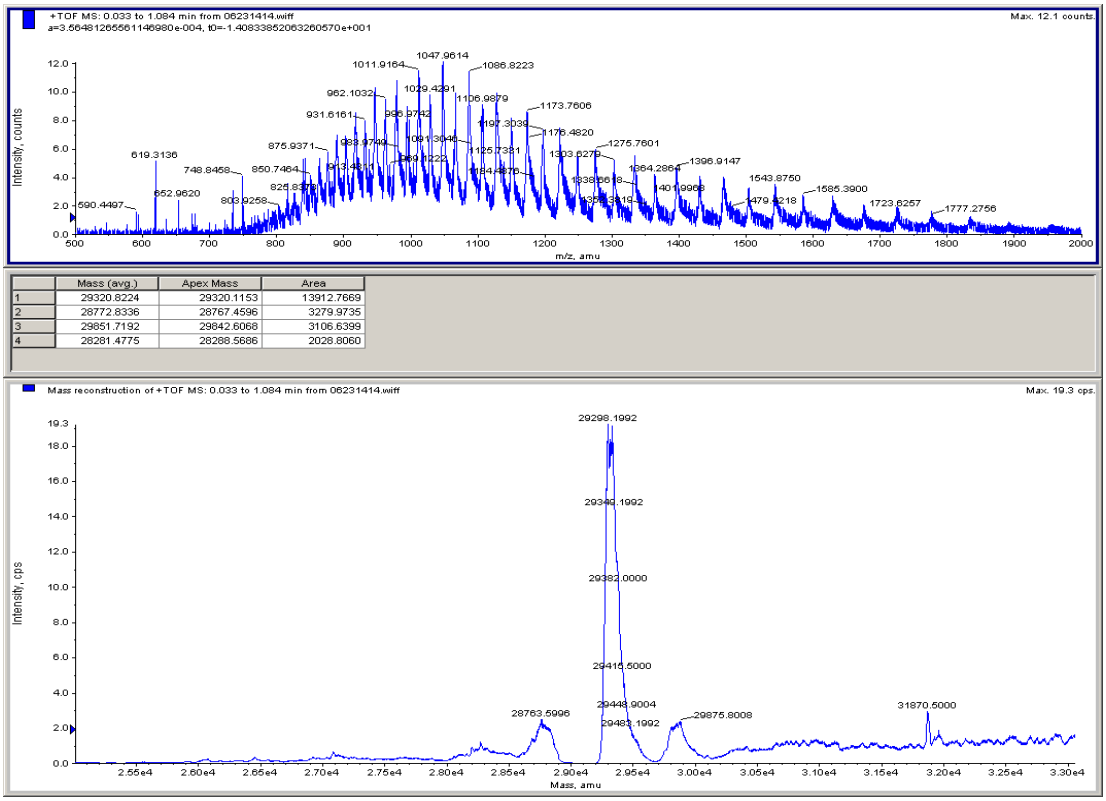


Figure 4.5. ESI-MS data for full-length Cys(Ala)$_5$-RCHO-sfGFP. Calculated mass: 29256 Da. Observed mass: 29298 Da (M + Acetyl).

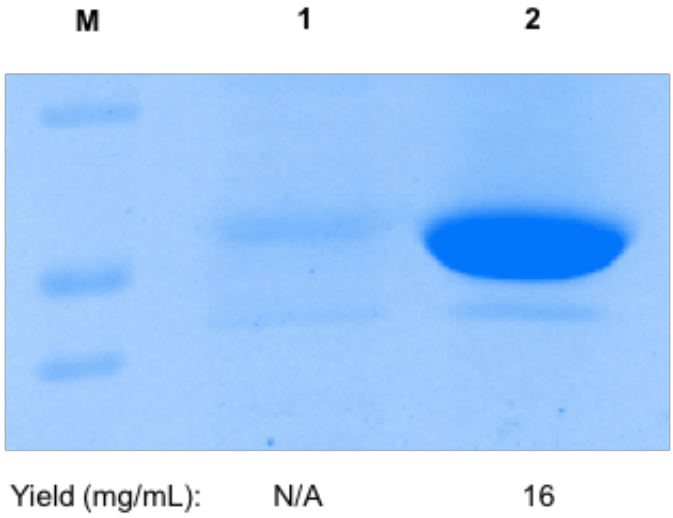


Figure 4.6. SDS-PAGE analysis of purified Cys(Ala)$_5$-RCHO-sfGFP stained with Coomassie brilliant blue. Lane 1: no amino acid supplemented. Lane 2: 2 mM 3-formyl phenylalanine.

N-Terminal Cysteine Labeling

A sample of Cys(Ala)$_5$-RCHO-sfGFP (100 µL, 8 µM) was incubated with TEV protease (final concentration: 0.8 µM) overnight (16 h) at 4 ºC. Then, a 20 µL aliquot was taken and treated with FITC-CBT for three hours. The reaction mixture was treated with 5 µL SDS-PAGE loading buffer, and the sample was boiled at 100 ºC for 10 min, then subjected to SDS-PAGE analysis. Prior to Coomassie staining, the gel was imaged to detect fluorescence from protein labeling.

pADLg3-Cys(NNK)$_7$-TAG Primers

pADLg3-Cys(NNK)$_7$-TAG Fwd:

5’ – TAA CAT GGC ATG CNN KNN KNN KNN KNN KNN KNN KTA GGC GGC GAA AGC GGC – 3’

PelB-NcoI-Rev:

5’-CATGCCATGGCCGGCTGGGCCGC-3’

pADLg3-Cys(NNK)$_7$TAG Sequence

DNA Sequence

atgaaatacctattgcctacggcggccgctggattgttattactcgcggcccagccggccatggcatgcnnknnknnknnkn

nknnknnktaggcggcgaaagcggccggcccgggaggccaaggcggtggttctgagggtggtggctccctcgagggcg

cgccagccgaaactgttgaaagttgtttagcaaaacctcatacagaaaattcatttactaacgtctggaaagacgacaaaacttta

gatcgttacgctaactatgagggctgtctgtggaatgctacaggcgttgtggtttgtactggtgacgaaactcagtgttacggtac

atgggttcctattgggcttgctatccctgaaaatgagggtggtggctctgagggtggcggttctgagggtggcggttctgagggt

ggcggtactaaacctcctgagtacggtgatacacctattccgggctatacttatatcaaccctctcgacggcacttatccgcctgg

tactgagcaaaaccccgctaatcctaatccttctcttgaggagtctcagcctcttaatactttcatgtttcagaataataggttccgaa

ataggcagggtgcattaactgtttatacgggcactgttactcaaggcactgaccccgttaaaacttattaccagtacactcctgtat

catcaaaagccatgtatgacgcttactggaacggtaaattcagagactgcgctttccattctggctttaatgaggatccattcgtttg

tgaatatcaaggccaatcgtctgacctgcctcaacctcctgtcaatgctggcggcggctctggtggtggttctggtggcggctct

gagggtggcggctctgagggtggcggttctgagggtggcggctctgagggtggcggttccggtggcggctccggttccggt

gattttgattatgaaaaaatggcaaacgctaataagggggctatgaccgaaaatgccgatgaaaacgcgctacagtctgacgct

aaaggcaaacttgattctgtcgctactgattacggtgctgctatcgatggtttcattggtgacgtttccggccttgctaatggtaatg

gtgctactggtgattttgctggctctaattcccaaatggctcaagtcggtgacggtgataattcacctttaatgaataatttccgtcaa

tatttaccttctttgcctcagtcggttgaatgtcgcccttatgtctttggcgctggtaaaccatatgaattttctattgattgtgacaaaat

aaacttattccgtggtgtctttgcgtttcttttatatgttgccacctttatgtatgtattttcgacgtttgctaacatactgcgtaataagga

gtcttaa

"nnk" denotes a randomized codon (n = A, T, C, or G; k = G or T). The amber stop codon is highlighted in red.

Peptide Sequence

M K Y L L P T A A A G L L L L A A Q P A M A C J J J J J J J **X** A A K A A G P G G Q G G G S E G G G S L E G A P A E T V E S C L A K P H T E N S F T N V W K D D K T L D R Y A N Y E G C L W N A T G V V V C T G D E T Q C Y G T W V P I G L A I P E N E G G G S E G G G S E G G G S E G G G T K P P E Y G D T P I P G Y T Y I N P L D G T Y P P G T E Q N P A N P N P S L E E S Q P L N T F M F Q N N R F R N R Q G A L T V Y T G T V T Q G T D P V K T Y Y Q Y T P V S S K A M Y D A Y W N G K F R D C A F H S G F N E D P F V C E Y Q G Q S S D L P Q P P V N A G G G S G G G S G G G S E G G G S E G G G S E G G G S E G G G S G G G S G S G D F D Y E K M A N A N K G A M T E N A D E N A L Q S D A K G K L D S V A T D Y G A A I D G F I G D V S G L A N G N G A T G D F A G S N S Q M A Q V G D G D N S P L M N N F R Q Y L P S L P Q S V E C R P Y V F G A G K P Y E F S I D C D K I N L F R G V F A F L L Y V A T F M Y V F S T F A N I L R N K E S

Where "J" denotes any of the twenty canonical amino acids, and **X** denotes a non-canonical amino acid.

pADLg3-Cys(NNK)$_7$-TAG Library Construction

Ten 50 µL scale PCR reactions were set up using Phusion DNA polymerase and the pADLg3-Cys(NNK)$_7$-TAG primers mentioned above. After PCR, each reaction was subjected to DpnI digestion for 2 h at 37 ºC and purified via spin column. The purified product was then subjected to NcoI digestion for 22 h at 37 ºC, followed by gel

purification. From here, the purified, NcoI-digested DNA was subjected to T4 ligation for 16 h at 16 ºC.

The ligated product from the above reaction was used to transform freshly prepared electrocompetent Top10 cells. To an ice-cold 0.6 mL tube was added 120 ng of pADLg3-Cys$(NNK)_7$-TAG (27.4 ng/µL), and a suspension of Top10 cells was slowly added by gently swirling the pipette while dispensing (100 µL total). From here, the suspension was transferred to a clean electroporation cuvette with a 2 mm gap. This procedure was repeated seven times. After each electroporation, cells were recovered by quickly adding 900 µL of pre-warmed (37 ºC) 2YT, and this 1 mL culture was recovered for 1 h at 37 ºC. Then, the individual recovered cultures were pooled and grown in 500 mL of 2YT supplemented with Amp (100 µL/mL) overnight. Finally, cells were harvested by centrifugation for 20 min at 4,000 rpm, followed by resuspension in 50 mL 20% glycerol/2YT supplemented with Amp (100 µL/mL).

5. THE COPPER-CATALYZED AZIDE-ALKYNE CYCLOADDITION "CLICK" REACTION REQUIRES TWO DISTINCT MONONUCLEAR COPPER (I)-CHELATING SPECIES

5.1. Introduction

The copper-catalyzed azide-alkyne cycloaddition (CuAAC) click reaction was first reported by Sharpless in 2002 and has been a boon to numerous fields, a fact underscored by thousands of citations the original reports have amassed.[32,36,180] The widespread use of this reaction can be attributed to its chemoselectivity and reactivity under a variety of conditions, and it is robust enough to proceed in complex media such as biological systems. The advent of CuAAC has greatly facilitated endeavors such as the bioorthogonal labeling of biomacromolecules,[39,181,182] the selective modification of carbon allotropes,[35,38,183-185] and the rapid assembly of combinatorial libraries.[37]

Despite numerous applications, the exact mechanism of CuAAC is still a contentious issue, though numerous mechanistic investigations have been reported[186-192] and a thorough review has recently been published.[193] Early reports highlighted the importance of the copper acetylide intermediate, and the involvement of copper(I) species in the catalytic cycle. The fact that the reaction is second-order with respect to copper(I) has led to postulation that two copper(I) atoms are involved in this process, either through dinuclear acetylide-based complexes or distinct chelation events for azide and alkyne. A recent report by Fokin[191] has shown the need for a two-copper(I) process, and an ESI-MS study by De Angelis[192] has reported direct observation of a dinuclear copper acetylide species. Most recently, a study by Bertrand[194] utilized stable N-

heterocyclic carbenes to isolate a dinuclear acetylide intermediate that was capable of undergoing CuAAC. Taken together, these publications remove some ambiguity regarding the nuclearity of CuAAC, and the most recent mechanistic proposal is shown in Figure **5.1.**

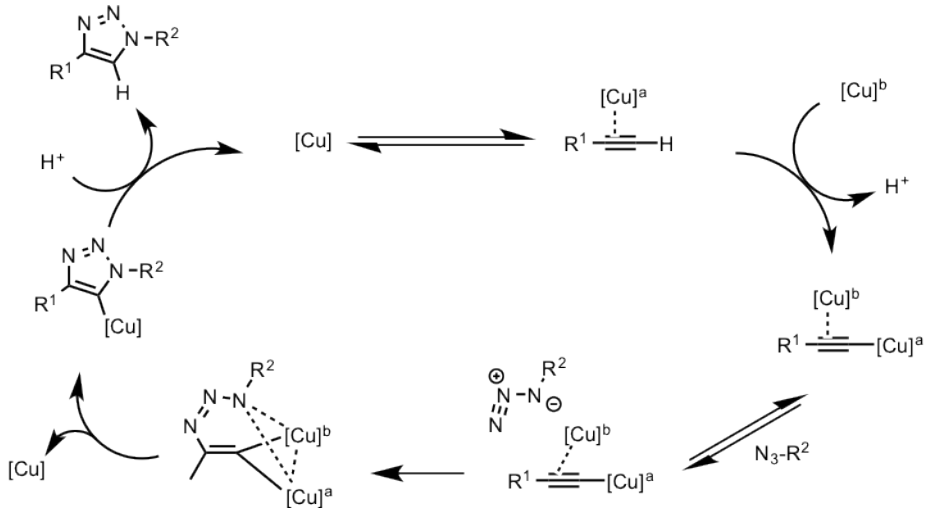


Figure 5.1. A proposed CuAAC mechanism by Fokin. Note the dinuclear copper intermediates in the bottom half of the mechanism.

The reports mentioned above assert that a dinuclear copper acetylide is crucial for reactivity to occur, a species activated by both σ- and π-complexation of the alkyne starting material. According to this model, azide would subsequently interact with this dinuclear intermediate to undergo cycloaddition. However, isolation of an intermediate

does not necessarily mean that it is mechanistically relevant, as it is possible for said intermediate to undergo rearrangement or decomposition to intermediates that do have significant roles in catalysis. Further, there may be dynamic and unstable short-lived events that are crucial to successful completion of the catalytic cycle. One of these events, copper-azide chelation, has previously been postulated as an important step in the reaction mechanism,[189,190] a possibility highlighted by several reports of CuAAC enhancement with copper-chelating azides.[39,40,182,195] Although current models acknowledge azide-copper chelation, its exact role and placement in the reaction mechanism has not been unequivocally demonstrated in light of modern studies. Indeed, most recent reports have mainly focused on the nature of alkyne-copper interactions, with minimal scrutiny given to azide-copper interactions.

Despite these advances in mechanistic understanding of CuAAC, some critical questions remain unanswered. How exactly are two coppers contributing to catalysis? Is dinuclear copper-alkyne binding a mechanistically relevant intermediate, or is it the product of a side reaction? In what manner does the azide interact with copper(I) and the alkyne? What exactly is the order of elementary chemical steps that lead to product formation? The dynamic nature of copper species in solution has made these questions challenging to answer, but we postulated that at least some of this information could be extracted via careful kinetic analysis.

5.2. Results and Discussion

To probe the reaction mechanism of CuAAC, we utilized two non-fluorescent coumarin dyes that have been previously reported.[40,196] The fluorogenic product

formation from CuAAC reactions of these two dyes allows for convenient monitoring of reaction trajectories via fluorescence. We initiated our investigations by using fluorescence-generating azido-coumarin **1** as the limiting reagent (Figure **5.2.A**), and ran the reaction under pseudo-first order conditions with largely excess propargyl alcohol in 100 mM potassium phosphate buffer. The other reagents such as $CuSO_4$, the copper(I) ligand BTTAA, and the reducing reagent sodium ascorbate were kept constant, and the exact concentrations used were those reported in a paper detailing the optimization of CuAAC.[34]

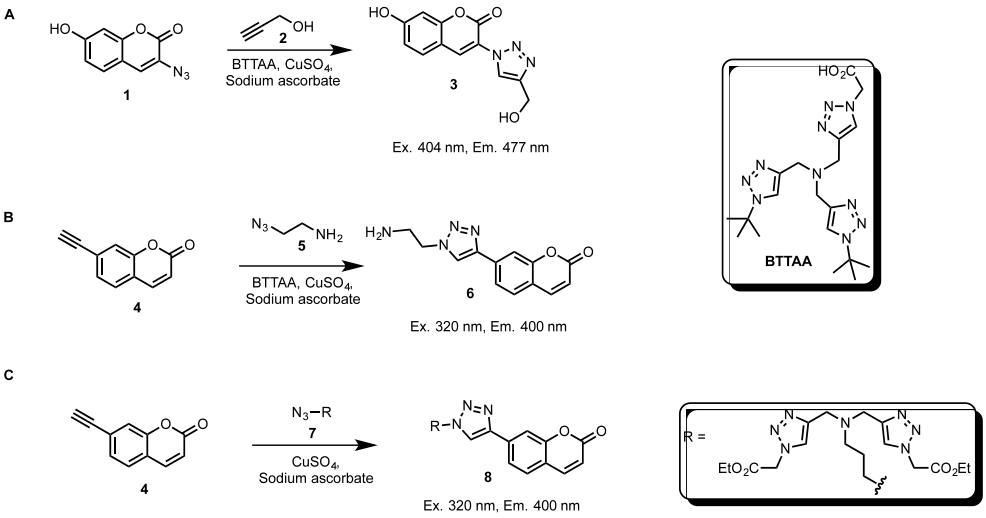


Figure 5.2. The model reactions used in this study. All reactions were carried out in 0.1 M potassium phosphate buffer, pH 7.4 with 5% DMSO.

Under these conditions, if there is an alkyne-copper(I) complex formation that is rate limiting or occurs prior to the rate-limiting step, then the reaction kinetics should show dependence based on alkyne concentration. The collected data of fluorescence intensity vs. time fit perfectly to the one phase exponential product formation model, allowing for readily determined observed reaction rate constants. However, there were no significant changes in observed rate constants over a number of different alkyne concentrations, indicating that the copper(I)-alkyne interaction, if there is any, and all other interactions involving alkyne occur after the rate-limiting step under the tested conditions (Figure **5.3**). This observation was an unexpected result since it is in conflict with the model shown in Figure **1** and most of other popular models of CuAAC.[191,192,194]

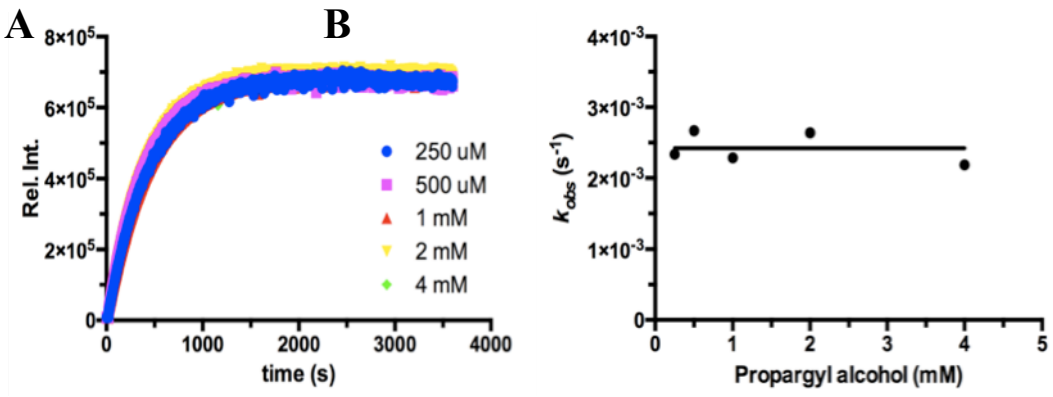


Figure 5.3. CuAAC reactions of azido-coumarin in excessive propargyl alcohol. (**A**) The one phase exponential increase of the product fluorescence at different propargyl alcohol concentrations. Data were fit to the equation $I_t = I_0 + I_C \, x \, 1 - e^{(-kobs \, x \, t)}$ where I_0 is the background fluorescent intensity, I_t the fluorescent intensity at a specified time, I_c the fluorescent intensity change when azido-coumarin is fully changed to its corresponding fluorescent final product, and k_{obs} the observed apparent rate constant. (**B**) Observed apparent rate constant k_{obs} vs. propargyl alcohol concentration (the data were fit to a flat line). Conditions: 50 μM azido-coumarin, 0.25-4 mM propargyl alcohol, 2.5 mM sodium ascorbate, 300 μM BTTAA, 50 μM $CuSO_4$.

We went to further probe the dependence of the reaction rate constant on the copper(I) concentration under conditions of largely excess propargyl alcohol. Using almost identical conditions as above with a fix concentration of propargyl alcohol at 1 mM, but with variable copper concentrations, we observed a linear dependence of observed rate constant on copper(I) concentration (Figure **5.4**). These results imply that only one stoichiometric equivalent of copper(I) is involved in the rate-determining step when alkyne is largely excessive, and that this process is much slower than steps involving alkyne. If any alkyne-involving step was rate limiting, or occurred prior to the

rate-limiting step, there would also be observable alkyne dependence, though that is not the case.

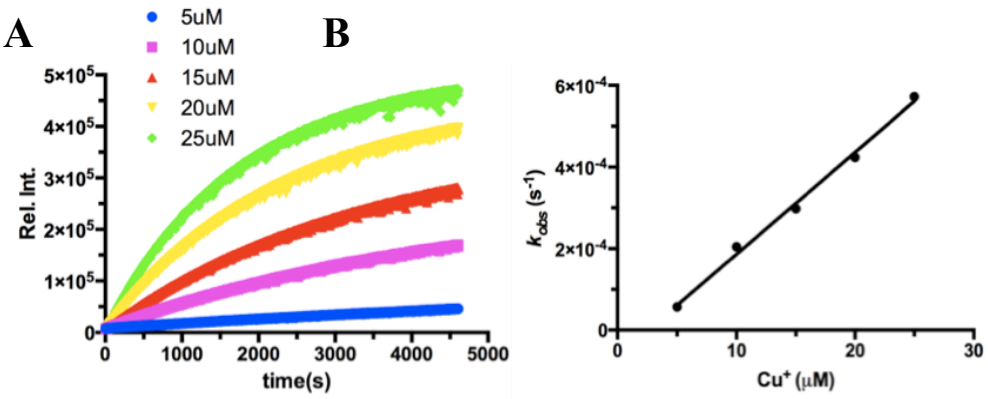


Figure 5.4. CuAAC reactions of azido-coumarin in excessive propargyl alcohol but various copper concentrations. (**A**) The collected data of product fluorescence intensity vs. time at different copper concentrations. Data were fit to the equation $I_t = I_0 + I_C \, x \, 1 - e^{(-kobs \, x \, t)}$ to obtain apparent observed rate constants at different conditions. (**B**) Observed apparent rate constant k_{obs} vs. copper concentration (data were fit to a linear curve). Conditions: 50 μM azido-coumarin, 1 mM propargyl alcohol, 2.5 mM sodium ascorbate, 300 + **x** μM BTTAA (**x** = [$CuSO_4$]), 5-25 μM $CuSO_4$.

Based on the above data, a mechanism of CuAAC under these conditions is proposed in Figure **5.5**. Under the conditions described above, the azide is the limiting reagent. Its one phase exponential decay, manifested by the one phase exponential increase of the product fluorescence, implies involvement of an azide-dependent process that must be

rate limiting. Our data also demonstrates that interaction with one copper(I) atom is important for this slow process, presumably via chelation to azide. Subsequent steps must involve the alkyne, and two possibilities exist: 1) a single copper(I) atom recruits both azide and alkyne reagents, or 2) another copper(I) atom activates the alkyne for cycloaddition with azide. Given that most of studies pointed to two copper(I) atoms in the catalysis of CuAAC, the latter seems most likely.

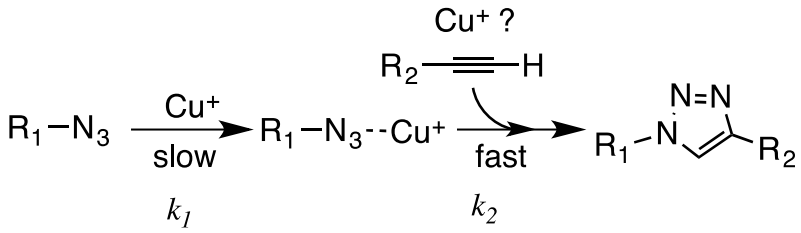


Figure 5.5. Limiting amounts of azide and excess amounts of alkyne lead to rate-limiting azide-copper(I) chelation. The observed apparent rate constant is determined by the equation $k_{obs} = k_1 \times [Cu^+]$.

In order to probe the alkyne involvement processes in CuAAC, we employed limiting amounts of a non-fluorogenic alkyne-coumarin **4** with excess amounts of azide **5**. These conditions effectively speed up the azide-copper(I)chelation process as well as slow down the alkyne-dependent steps and therefore potentially ensure

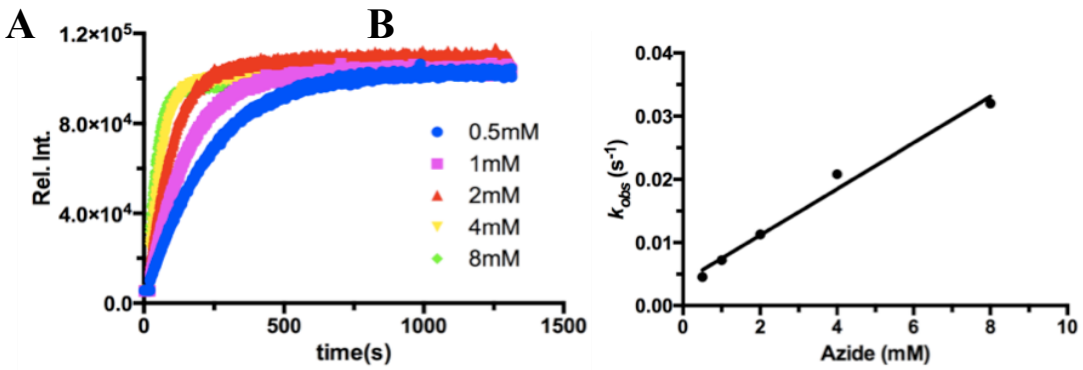


Figure 5.6. CuAAC reactions of alkyne-coumarin in excessive 2-azidoethanamine. (**A**) The collected data of product fluorescence intensity vs. time at different 2-azidoethanamine concentrations. Data were fit to the equation $I_t = I_0 + I_C \, x \, 1 - e^{(-kobs \, x \, t)}$ to obtain observed apparent rate constants at different conditions. (**B**) Observed apparent rate constant k_{obs} vs. 2-azidoethanamine concentration (data were fit to a linear curve). Conditions: 50 μM alkyne-coumarin, 0.5-8 mM 2-azidoethanamine, 2.5 mM sodium ascorbate, 300 μM BTTAA, 50 μM $CuSO_4$.

that alkyne-based processes are rate limiting and observable. As before, conditions were based on the optimized concentrations reported previously,[34] with micromolar amounts of **4** and millimolar amounts of 2-azidoethanamine. Under these conditions, the determined reaction rate constant shows linear dependence on azide concentration (Figure **5.6**), indicating that the azide-copper(I) chelation occurs before further reaction with alkyne, a result that agrees well with observations using azido-coumarin **1**. We also carried out the copper dependence study under conditions of largely excessive azide and limiting alkyne. The determined observed apparent rate constant vs. copper(I) concentration displayed a clearly second-order dependence (Figure **5.7**), an observation

made by others and one that illustrates the importance of two distinct copper(I) atoms in the catalysis.

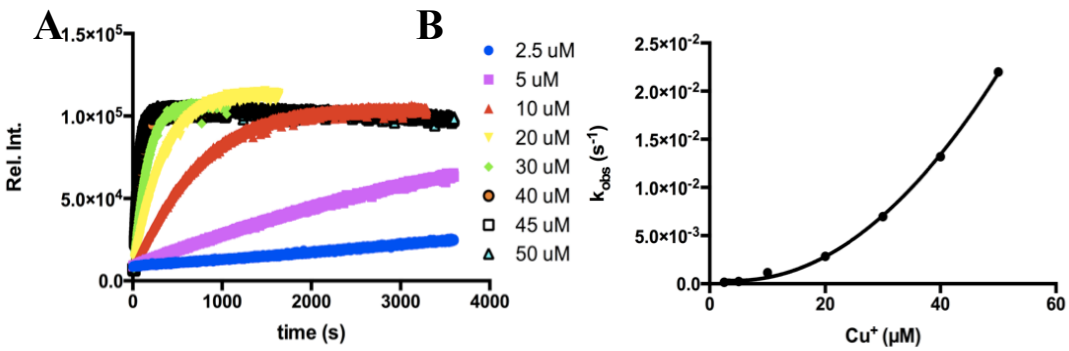


Figure 5.7. CuAAC reactions of alkyne-coumarin in excessive 2-azidoethanamine but various copper concentrations. (**A**) The collected data of product fluorescence intensity vs. time at different copper concentrations. Data were fit to the equation $I_t = I_0 + I_C \, x \, 1 - e^{(-kobs \, x \, t)}$ to obtain observed apparent rate constants at different conditions. (**B**) Observed apparent rate constant k_{obs} vs. copper concentration (data were fit to a second-order model equation $y = Ax^2 + B$). Conditions: 50 μM alkyne-coumarin, 10 mM 2-azidoethanamine, 2.5 mM sodium ascorbate, 300 + **x** μM BTTAA (**x** = [$CuSO_4$]), 2.5-50 μM $CuSO_4$.

Based on observations described above, we propose a CuAAC mechanism in conditions of largely excessive azide and limiting alkyne shown in Figure **5.8**. Under these conditions, the excessive azide drives the fast formation of an azide-copper(I) complex that quickly reaches equilibrium with free azide and copper(I). This azide-copper(I) complex then involves a relatively slow reaction with alkyne and a copper(I)

atom to give the final cycloaddition product. Under this mechanism, the determined product formation displays a first-order dependence on the azide concentration and a second-order dependence on the copper(I) concentration.

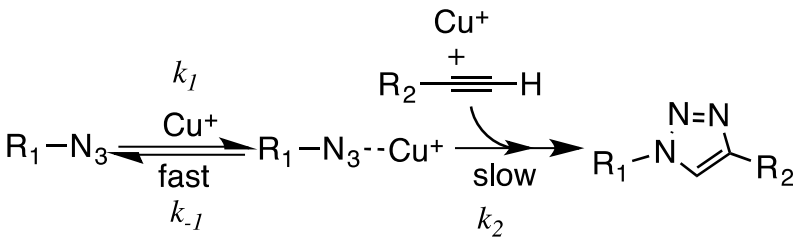


Figure 5.8. Limiting amounts of alkyne-coumarin **4** results in rate-limiting processes involving alkyne. The observed apparent rate constant is determined by the equation $k_{obs} = k_2 \, x \, [Cu^+] \, x \, [R_1\text{-}N_3\text{--}Cu^+]$ that can be further derived to be the equation $k_{obs} = [(k_1 \, x \, k_2)/k_{-1}] x \, [Cu^+]^2 \, x \, [R_1\text{-}N_3]$. When $k_1 >> k_{-1}$, both free $R_1\text{-}N_3$ and Cu^+ concentrations are essentially not varied due to their chelation.

The kinetic data taken thus far have narrowed down the mechanism to two possible pathways shown in Figure **5.9**. The pathway shown in blue involves a mononuclear transition state arrangement in which both azide and alkyne chelate to a single copper(I) atom. Subsequent interaction with a second copper(I) atom would somehow trigger formation of the final product, perhaps via π-complexation with alkyne; such an intermediate would be analogous to the most recent mechanistic proposal.[191,192,194] The

second pathway (bottom, shown in red) involves two distinct copper(I) species, one chelated to azide and another interacting with alkyne, most likely as the sigma complex. Both copper(I)-chelating species would react together to form the triazole product.

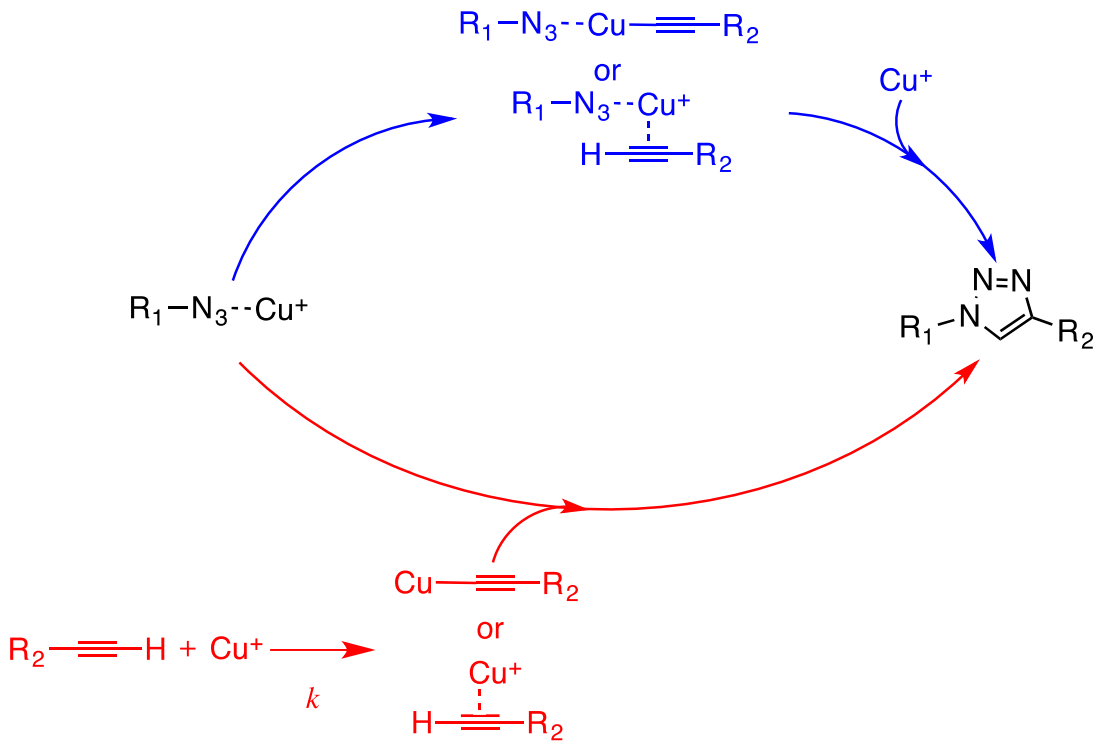


Figure 5.9. Depiction of two mechanistic possibilities based on kinetic data shown above.

In order to distinguish between the two possibilities, a copper(I)-chelating azide **7** was synthesized and is analogous to compounds reported recently.[40] A number of reports

have been published describing the rapid rate enhancement of CuAAC when using copper-chelating azides,[39,40] and we suspected that such a compound would facilitate kinetic investigations. By utilizing a copper(I)-chelating azide, rapid formation of a copper(I)-azide complex should occur, making other processes rate limiting. Using stopped-flow kinetics and a limiting amount of the chelating azide **7**, a one-phase exponential product formation is expected if the mechanism in blue is correct (i.e. $[R_1\text{-}N_3]_t = [R_1\text{-}N_3]_0(1 - e^{-kt})$). However, under these conditions we observed a linear formation of product followed by a flat line that was not the typical, exponential curve for product formation (Figure **5.10.A**). This data indicates that processes involving azide are not rate limiting, and that the rate of the product formation is determined by another slow step; based on process of elimination, this event must be copper(I)-alkyne complexation shown in red in Figure **5.9**. Under the mechanism in red of Figure **5.9**, where **7** is limiting and the copper(I)-alkyne complex formation is much slower than the following reaction

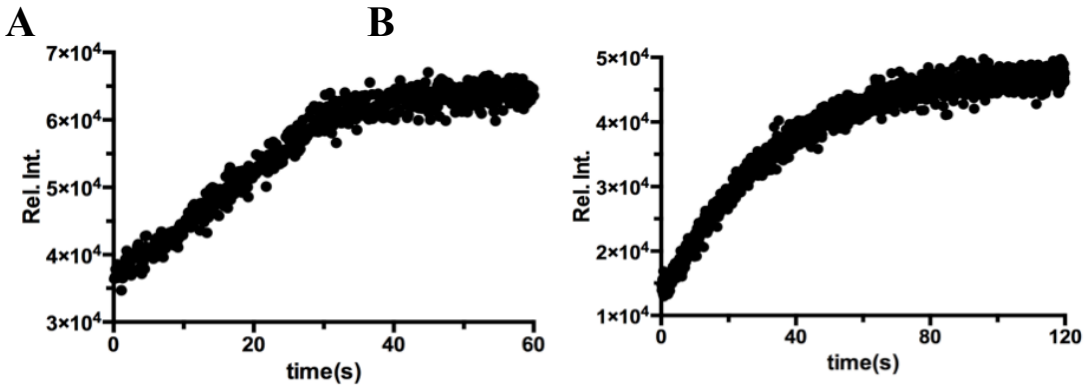


Figure 5.10. Stopped-flow kinetics of CuAAC between alkyne-coumarin **4** and a copper-chelating azide **7**. (**A**) The product fluorescence intensity vs. time when **7** was limiting. Conditions: 20 μM **7**, 200 μM alkyne-coumarin, 20 μM $CuSO_4$, 2.5 mM sodium ascorbate. (**B**) The product fluorescence intensity vs. time when **4** was limiting. Conditions: 20 μM alkyne-coumarin, 200 μM **7**, 20 μM $CuSO_4$, 2.5 mM sodium ascorbate.

with the azide-copper(I) complex, the rate of consumption of **7** (or the rate of product formation) is determined by copper(I)-alkyne complexation with a constant rate defined by the equation $v = k\ x\ [alkyne]\ x\ [Cu+]$ in which both concentrations of alkyne and copper(I) do not vary during reaction. We also tested reactions under conditions of limiting alkyne and excessive **7**, and we once again observed the exponential increase in product formation (Figure **5.10.B**). Under these conditions, the product formation manifested by the alkyne consumption follows a one-phase exponential increase model. Taken together, these experiments lead to rejection of the pathway shown in blue, and provide strong evidence to suggest the mechanism shown in red of Figure **5.9** is correct and the real cycloaddition reaction is much faster than the alkyne-copper(I) complexation. A revised mechanism is shown in Figure **5.11**. In this mechanism, both alkyne and azide chelate copper(I) first and then react rapidly to form the final

cycloaddition product. The cycloaddition process is much faster than the two complexation steps, leading to all detected reaction rates in this study being related to copper(I)-azide and copper(I)-alkyne complexation.

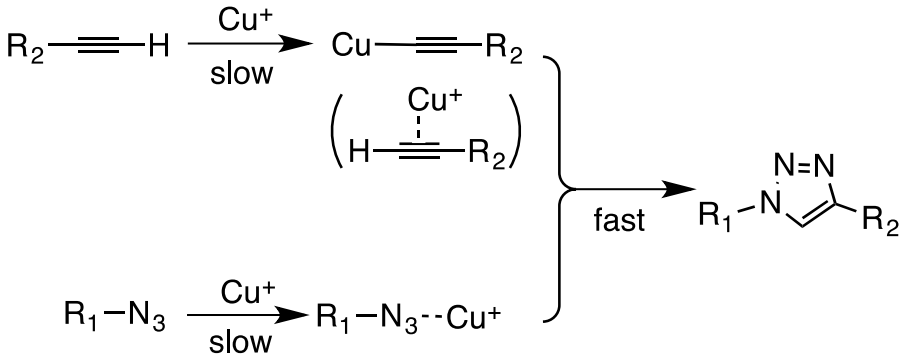


Figure 5.11. A refined mechanism based on the results of this study.

5.3. Concluding Remarks

The most recent mechanistic studies have shed light on the nuclearity of CuAAC, and the consensus currently points to formation of an acetylide intermediate activated via σ- and π-complexation. After formation of this dinuclear intermediate, azide chelation is postulated to occur, followed by cycloaddition. However, if this step occurred as hypothesized, our results would have differed in significant ways. For

example, in situations where azido-coumarin is used in the presence of excess propargyl alcohol, we would have observed a second-order dependence on the copper(I) concentration that is typically seen for this reaction, in addition to linear dependence on the propargyl alcohol concentration. These hypothetical results would have supported rate-limiting formation of the dinuclear acetylide intermediate. The observation that the rate is independent of alkyne and first-order for copper leads to rejection of this order of events. Hypothetically, if azide chelation occurred after the rate-limiting step, we should observe zero order dependence on azide concentration. However, our observation that this reaction is first order with excess azide dismisses this idea.

It is important to recognize that the successful isolation of an intermediate does not necessarily indicate this intermediate to be mechanistically relevant in a direct way. It is entirely possible that such an intermediate undergoes decomposition or rearrangement in solution to reveal intermediates that are actually involved in the mechanism. In the case for CuAAC, stable dinuclear acetylides have been isolated and utilized for triazole formation [194]. Given the kinetic results in this paper, we propose that the π-bound copper from such an intermediate would have to dissociate from the acetylide to interact with azide, and the resulting complex would then undergo cycloaddition with copper acetylide. In other words, although the dinuclear copper acetylide can be isolated under certain conditions, it is most likely a non-productive intermediate that dissociates to generate mechanistically relevant species. Further, the successful isolation of this doubly activated alkyne required ligands that are electronically tuned to stabilize this

intermediate. Typically, copper-alkyne π-complexation is a weak interaction under standard conditions with conventional reagents.

Prior to this work, there have been no reports that scrutinize the currently accepted mechanism with detailed kinetic analysis, which can elucidate the order of events in such a dynamic system. Given the results of our kinetic experiments, the proposed dinuclear copper acetylide does not occur prior to azide chelation events. Rather than delegating azide chelation to a later step in the catalytic cycle, we propose that it is a distinct chelation event that requires one atom of copper(I), and the other copper(I) atom activates the alkyne, most likely through an acetylide intermediate. After formation of both copper complexes, they come together to form the triazole product (Figure **5.11**), possibly through π-interaction of the copper-azide complex with the σ-bound copper acetylide. Reports demonstrating the effectiveness of copper-chelating azides in CuAAC further emphasize the importance of azide chelation events in this reaction.

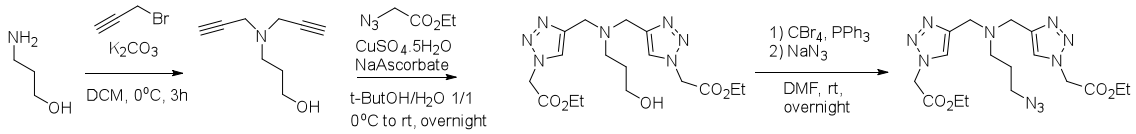
We believe the results presented herein will help researchers develop new, more efficient catalysts for such a widely used reaction. Copper-chelating azides have been successful at enhancing the reaction rate of CuAAC, but these were designed with previous mechanistic models in mind. With our current understanding of copper-catalyzed click chemistry, it may be possible to synthesize azides that further enhance the rate and efficiency of this venerable reaction. Through informed optimization of reagent and catalyst properties, it is expected that researchers will be successful in enhancing copper-catalyzed click chemistry even further than previously reported.

5.4. Detailed Experimental Protocols

Unless otherwise noted, all reagents and solvents were used as purchased without further purification. All reactions were performed under an inert atmosphere, using oven-dried glassware and anhydrous solvents. Azido-coumarin **1**, alkyne-coumarin **4**, and BTTAA were synthesized based on published literature procedures.[197-200]

An important caveat regarding CuAAC is that freshly prepared solutions of copper (ii) sulfate and sodium ascorbate are critical for success of the reaction, and efficacy of click chemistry diminishes with increased age of the stock solutions. Consequently, for the kinetic experiments described below, all measurements for a single plot were taken in a single day, using freshly prepared solutions of copper (ii) sulfate and sodium ascorbate. Enough stock solution was made for the anticipated number of reactions, in order to minimize experimental errors in preparing stock solutions that may lead to erroneous plot points. Data was collected using Felix GX software, and the raw data was transported to Kaleidagraph for data processing and curve fitting. For aesthetic reasons, the data shown in the main text was re-processed using Prism 6 software. The original data and curve fitting has been included in this document for the sake of transparency.

5.4.1. Organic Synthesis



9 **10** **7**

*3-(Di(prop-2-yn-1-yl)amino)propan-1-ol (**9**)*

A previously reported procedure has been adapted.[201] 3-Amino-1-propanol (1.5g, 20mmol) and K_2CO_3 (5.53g, 40mmol) was dissolved in 10 mL DCM and placed in an ice bath. Then, propargyl bromide (5.95g, 40mmol) in 3 mL DCM was added dropwise. After 3 hours, the reaction had reached completion, and the solvent was evaporated under reduced pressure. The crude material was directly subjected to column chromatography (DCM/EtOAc: 1/1) to afford **9** in 33% yield. ^{1}H NMR (D_2O, 300 MHz): δ 3.63 (t, 2H, *J*=6.5Hz), 3.47 (s, 4H), 2.66 (t, 2H, *J*=7.5Hz), 1.73 (p, 2H, *J*=7.1Hz). ^{13}C NMR (D_2O, 75 MHz): δ 78.21, 74.69, 59.83, 49.39, 41.35, 28.89.

*2,2'-(((3-Hydroxypropyl)azanediyl)bis(1H-1,2,3-triazole-4,1-diyl))diacetic Acid(**10**)*

A previously reported procedure has been adapted.[40] Compound **9** (317 mg, 2.1mmol), ethyl 2-azido acetate (520mg, 4.2mmol, prepared according to a previously published report),[202] and 5 mL tert butanol/water (1/1) were added to the reaction flask. Then, $CuSO_4.5H_2O$ (52 mg, 0.21mmol) was added. Finally, the reaction flask was

placed in an ice bath, and sodium ascorbate (416 mg, 2.1mmol) was added in small portions over 2 minutes, then the ice bath was removed and the reaction was allowed to run overnight. After the reaction had reached to completion, solvent was evaporated under reduced pressure and the crude material was directly subjected to column chromatography (EtOAc/MeOH: 8/2) to afford **10** in 88% yield. ^{1}H NMR (D_2O, 300 MHz): δ 8.02 (s, 2H), 5.37 (s, 4H), 4.25 (q, 4H, *J*=7.2Hz), 3.85 (s, 4H), 3.59 (t, 2H, *J*=6.3Hz), 2.59 (t, 2H, *J*=7.7Hz), 1.81 (p, 2H, , *J*=7.0Hz), 1.25 (t, 6H, *J*=7.7Hz). ^{13}C NMR (D_2O, 75 MHz): δ 168.58, 142.90, 126.61, 63.20, 60.05, 50.91, 49.60, 46.92, 28.29, 13.15.

*2,2'-(((3-Azidopropyl)azanediyl)bis(1H-1,2,3-triazole-4,1-diyl))diacetic Acid (**7**)*

Compound **10** (190 mg, 0.5mmol), carbon tetrabromide (249mg, 0.75mmol) and 2 mL anhydrous DMF were added to the reaction flask, and then the flask was placed in an ice bath. Triphenyl phosphine (197 mg, 0.75mmol) was added and mixture was stirred for 30 minutes on ice under inert atmosphere. After the allotted time had passed, one drop of water was added to the reaction mixture. Following this, NaN_3 (195mg, 3mmol) was added, then the ice bath was removed and the reaction was stirred overnight. After the reaction had reached completion, solvent was evaporated under reduced pressure and the crude material was directly subjected to column chromatography (EtOAc) to afford **7** in 49% yield. ^{1}H NMR (D_2O, 300 MHz): δ 8.01 (s, 2H), 5.37 (s, 4H), 4.27 (q, 4H, *J*=7.2Hz), 3.86 (s, 4H), 3.31 (t, 2H, *J*=6.5Hz), 2.56 (t, 2H, *J*=7.3Hz), 1.82 (p, 2H, *J*=7.1Hz), 1.26 (t, 6H, *J*=7.1Hz). ^{13}C NMR (CD_3OD, 75 MHz): δ 168.58, 145.75, 127.02, 63.32, 51.91, 51.30, 50.51, 27.73, 14.53

ESI Mass Analysis (m/z): Calculated: 435.2217; Found: 435.2225

5.4.2. General Procedure for Pseudo-First Order Reactions

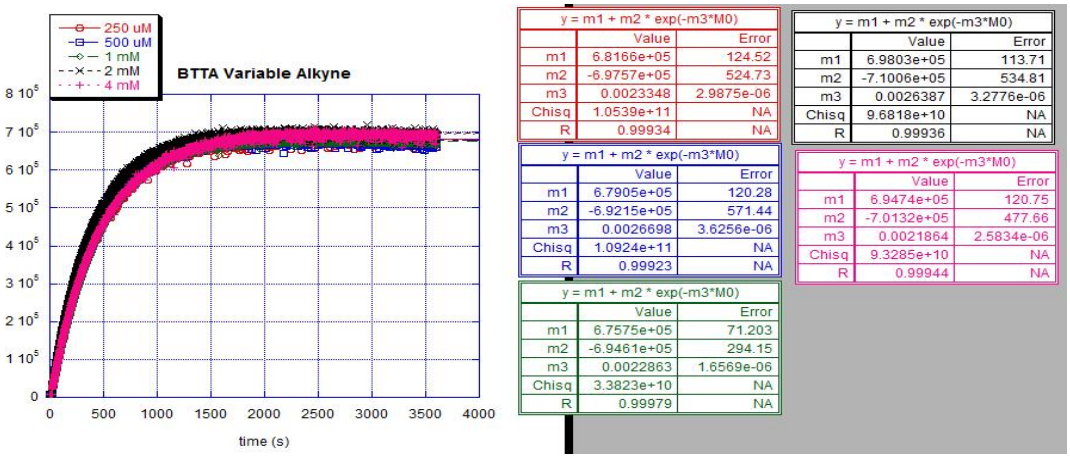
Azido-Coumarin

Table 5.1. Reagent table for pseudo-first order CuAAC with azido-coumarin.

Reagent	[Stock]	[Rxn]	Vol. (2000 μL total)
$CuSO_4$	25 mM	50 μM	4 μL
BTTAA	20 mM	300 μM	30 μL
DMSO	100%	5%	100 μL
KH_2PO_4	0.4 M	0.1 M	500 μL
Azido-coumarin	20 mM	50 μM	5 μL
Propargyl alcohol	1 M	1 mM	2 μL
Sodium ascorbate	25 mM	2.5 mM	200 μL
H_2O	N/A	N/A	1159 μL

Copper (ii) sulfate was added to an empty cuvette, followed by addition of BTTAA. The resulting solution was mixed well, followed by DMSO, potassium phosphate buffer, and water (in that order). Azido-coumarin was then added to the solution, followed by sodium ascorbate, and the resulting reaction mixture was placed in the fluorometer. Data acquisition was initiated for fifteen seconds, then acquisition was

paused, propargyl alcohol was added, and data acquisition was resumed. When altering the concentration of one reagent, care was taken to ensure the volume was identical in all cases (2000 μL) in order to minimize error, a precaution easily done by increasing or decreasing the volume of water necessary. For trials labeled "adjusted BTTAA", the actual BTTAA concentration was (300 + x) μM, where x = [$CuSO_4$].



y = m1 + m2 * exp(-m3*M0)	Value	Error
m1	6.8166e+05	124.52
m2	-6.9757e+05	524.73
m3	0.0023348	2.9875e-06
Chisq	1.0539e+11	NA
R	0.99934	NA

y = m1 + m2 * exp(-m3*M0)	Value	Error
m1	6.9803e+05	113.71
m2	-7.1006e+05	534.81
m3	0.0026387	3.2776e-06
Chisq	9.6818e+10	NA
R	0.99936	NA

y = m1 + m2 * exp(-m3*M0)	Value	Error
m1	6.7905e+05	120.28
m2	-6.9215e+05	571.44
m3	0.0026698	3.6256e-06
Chisq	1.0924e+11	NA
R	0.99923	NA

y = m1 + m2 * exp(-m3*M0)	Value	Error
m1	6.9474e+05	120.75
m2	-7.0132e+05	477.66
m3	0.0021864	2.5834e-06
Chisq	9.3285e+10	NA
R	0.99944	NA

y = m1 + m2 * exp(-m3*M0)	Value	Error
m1	6.7575e+05	71.203
m2	-6.9461e+05	294.15
m3	0.0022863	1.6569e-06
Chisq	3.3823e+10	NA
R	0.99979	NA

Figure 5.12. Reaction progress curves for azido-coumarin with excess propargyl alcohol.

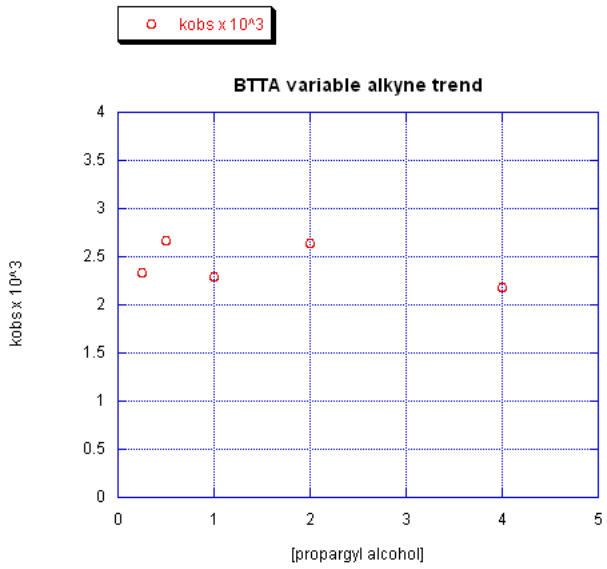


Figure 5.13. Reaction rate dependence on propargyl alcohol concentration.

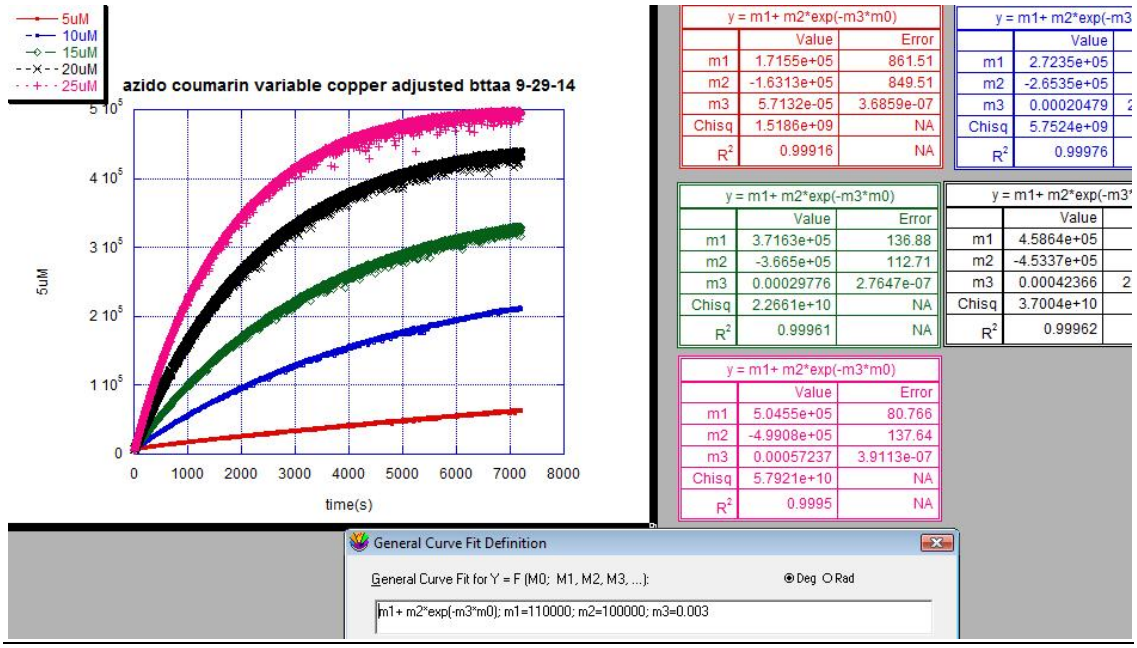


Figure 5.14. Reaction progress curves for azido-coumarin with excess propargyl alcohol and variable copper sulfate concentration.

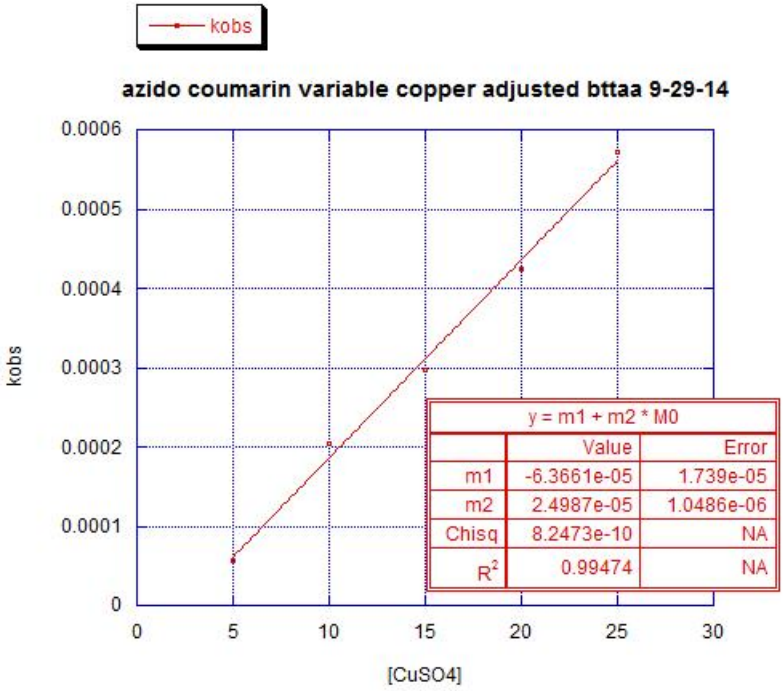


Figure 5.15. Reaction rate dependence on copper sulfate concentration.

Alkyne-Coumarin

Table 5.2. Reagent table for pseudo-first order CuAAC with alkyne-coumarin.

Reagent	[Stock]	[Rxn]	Vol. (2000 μL total)
$CuSO_4$	25 mM	50 μM	4 μL
BTTAA	20 mM	300 μM	30 μL
DMSO	100%	5%	100 μL
KH_2PO_4	0.4 M	0.1 M	500 μL
Alkynyl-coumarin	20 mM	50 μM	5 μL
2-azidoethanamine	1 M	1 mM	2 μL
Sodium ascorbate	25 mM	2.5 mM	200 μL
H_2O	N/A	N/A	1159 μL

The procedure for alkyne-coumarin was identical to the one described above, except alkyne-coumarin and 2-azidoethanamine was added instead of azido-coumarin and propargyl alcohol, respectively. Again, caution was taken to ensure that the reaction volume was identical in all cases. For trials labeled “adjusted BTTAA”, the concentration of BTTAA was equal to (300 + x) μM, where “x” is equal to the concentration of copper sulfate.

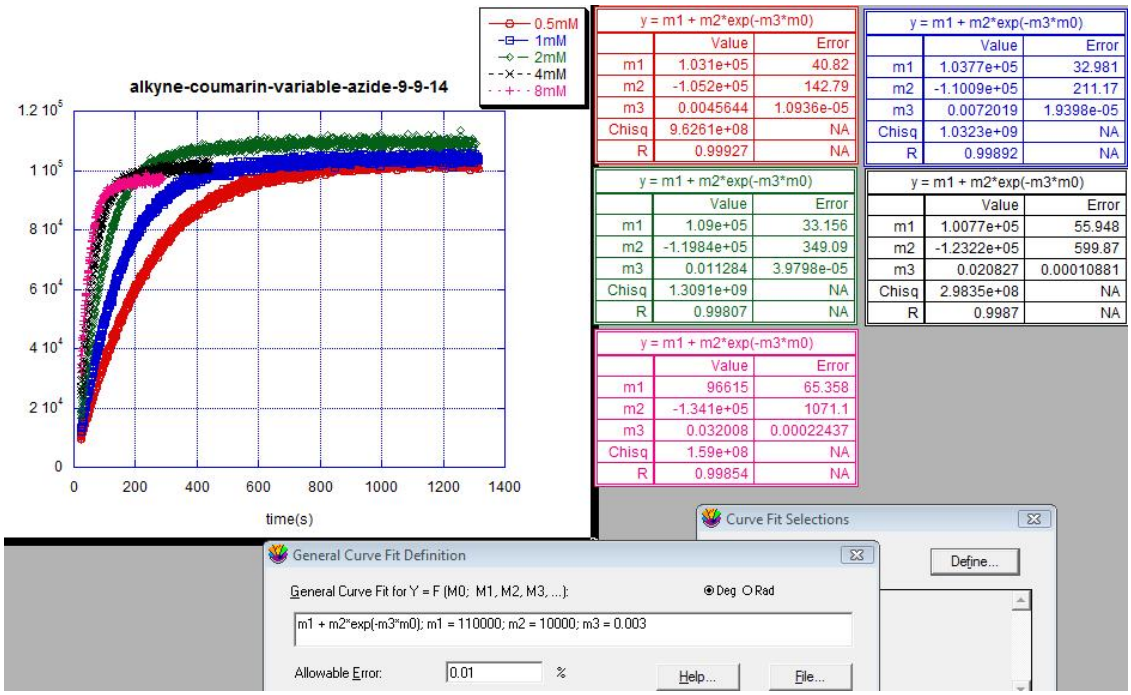


Figure 5.16. Reaction progress curves for alkyne-coumarin with excess 2-azidoethanamine.

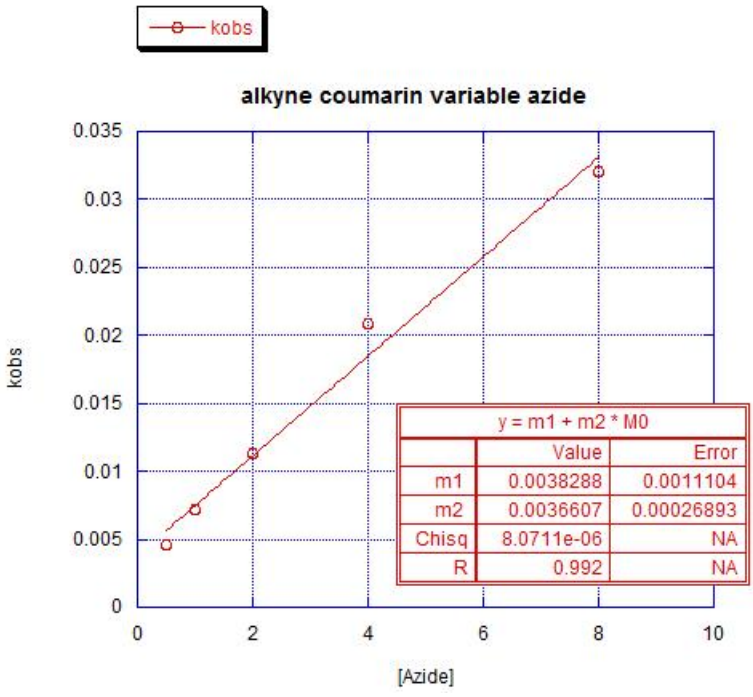


Figure 5.17. Reaction rate dependence on 2-azidoethanamine concentration.

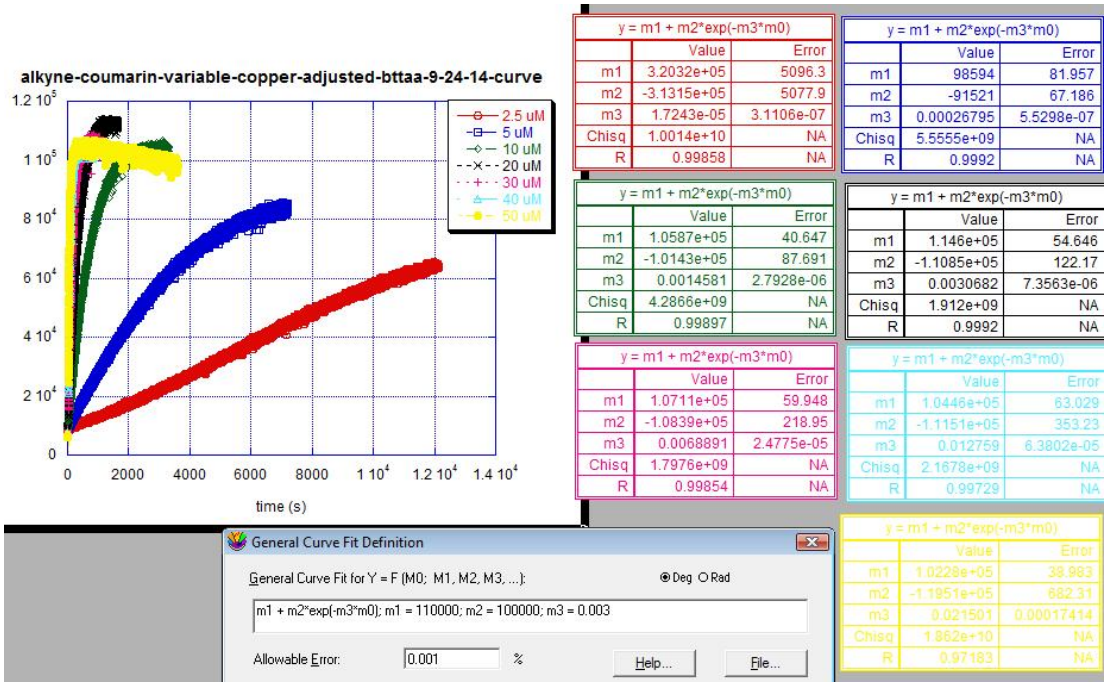


Figure 5.18. Reaction progress curves for alkyne-coumarin with excess 2-azidoethanamine and variable copper sulfate concentration.

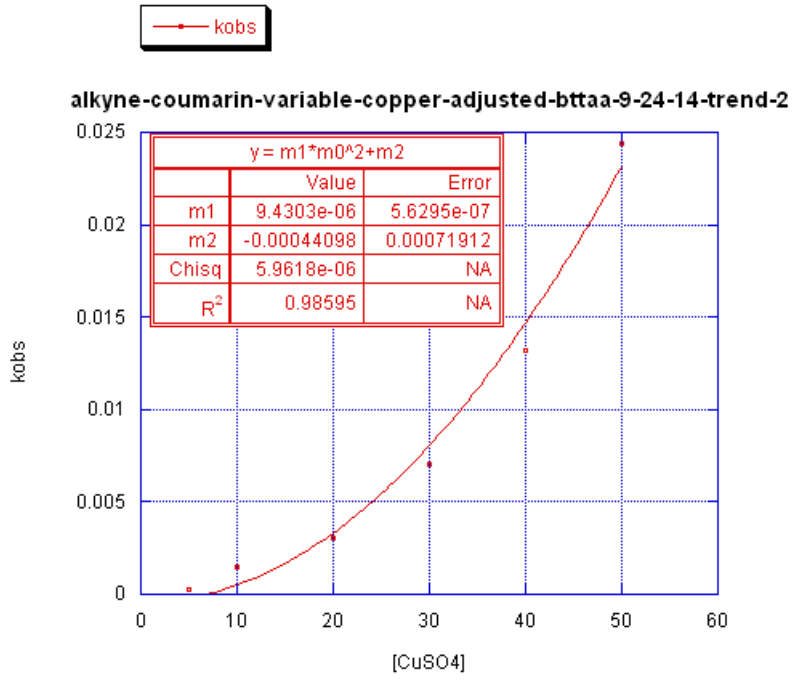


Figure 5.19. Reaction rate dependence on copper sulfate concentration.

5.4.3. General Procedure for Stopped-Flow Kinetics

To assess the kinetic parameters of the copper-chelating azide, two separate 1 mL solutions were prepared. To prepare solution **A**, reagents were added to a clean 1.5 mL Eppendorf tube in the following order: copper (ii) sulfate, azide **7**, DMSO, potassium phosphate buffer, water, and sodium ascorbate. Solution **B** was a 1 mL solution of alkyne-coumarin in water, prepared in a separate 1.5 mL Eppendorf tube. Both solutions were independently taken up using two 1 mL syringes and simultaneously injected into the stopped flow apparatus. Data acquisition was triggered to occur at the moment injection was complete.

Excess Azide

Table 5.3. Reagent table for stopped flow kinetics with excess azide **7**.

Reagent	[Stock]	[Rxn]	Soln. **A** (1000 μL)	Soln. **B** (1000 μL)
$CuSO_4$	25 mM	50 μM	4 μL	
DMSO	100%	5%	100 μL	
KH_2PO_4	0.4 M	0.1 M	500 μL	
Alkyne-coumarin	20 mM	20 μM		2 μL
Azide **7**	20 mM	200 μM	20 μL	
Sodium ascorbate	25 mM	2.5 mM	200 μL	
H_2O	N/A	N/A	176 μL	998 μL

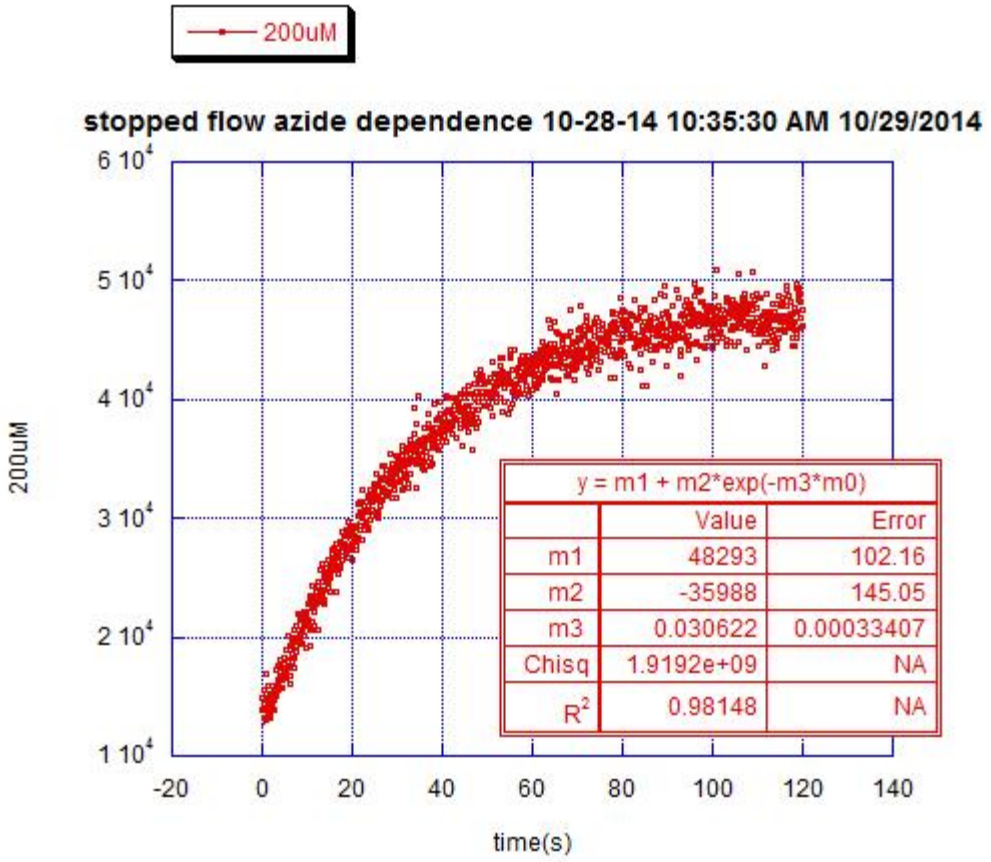


Figure 5.20. Product formation curve for excessive azide **7** and limiting amounts of alkyne-coumarin.

Excess Alkyne

Table 5.4. Reagent table for stopped flow kinetics with excess alkyne-coumarin.

Reagent	[Stock]	[Rxn]	Soln. **A** (1000 μL)	Soln. **B** (1000 μL)
$CuSO_4$	25 mM	50 μM	4 μL	
DMSO	100%	5%	100 μL	
KH_2PO_4	0.4 M	0.1 M	500 μL	
Alkyne-coumarin	20 mM	200 μM		20 μL
Azide **7**	20 mM	20 μM	2 μL	
Sodium ascorbate	25 mM	2.5 mM	200 μL	
H_2O	N/A	N/A	194 μL	980 μL

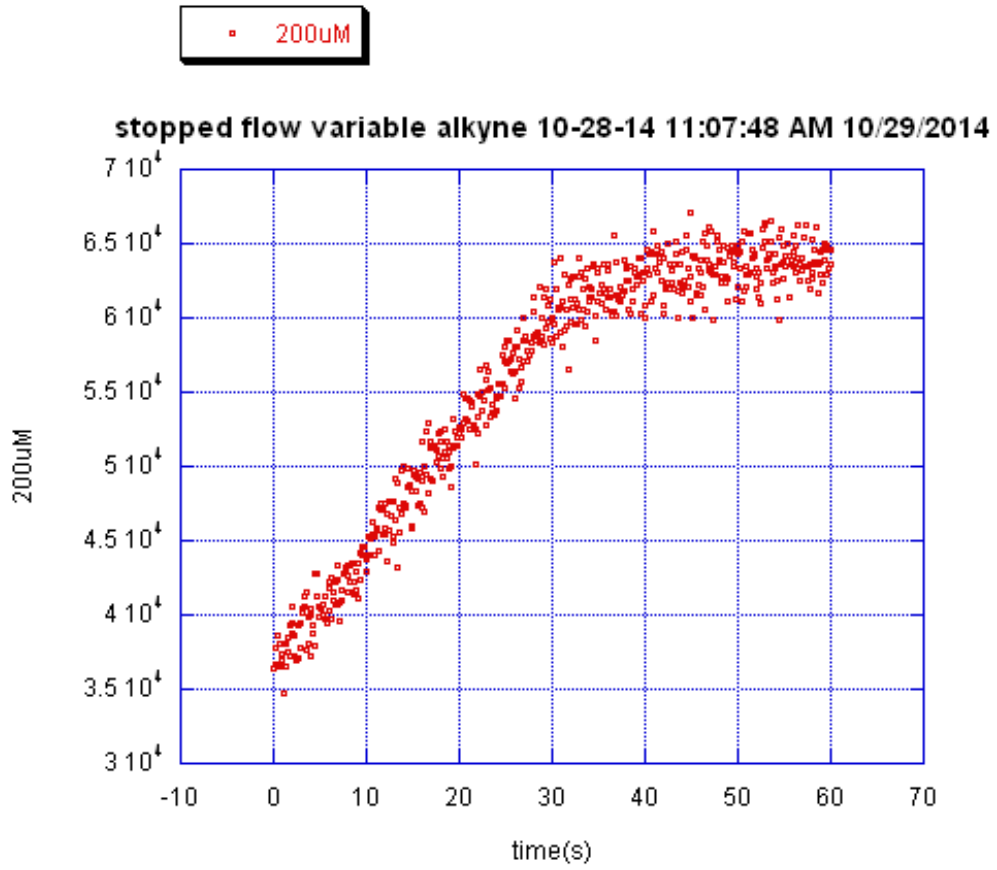


Figure 5.21. Product formation curve for limiting azide **7** and excessive amounts of alkyne-coumarin.

6. SUMMARY

The results presented herein have demonstrated the utility of non-canonical phenylalanine derivatives in chemical biology for probing biological systems. Using an expanded genetic code of long-chain, *meta*-substituted phenylalanine derivatives, the steric tolerances of PylRS (N346A/C348) were explored, and its ability to incorporate all of the long-chain substrates is a testament to its superb promiscuity, a desirable trait because a variety of amino acids can be installed with a single system, minimizing the need for evolving novel mutants for each substrate. This information has been utilized to synthesize sterically similar amino acids for incorporation with PylRS(N346A/C348). Moreover, it has been demonstrated that PylRS (N346A/C348A) can successfully incorporate 3-formyl-phenylalanine, an aldehyde-containing amino acid that is highly reactive at physiological pH, yet is robust enough to survive protein expression conditions. This marks the first time that aldehydes could be directly incorporated using amber suppression methodology; previous efforts utilized indirect methods of chemical or enzymatic post-translational modifications, and these approaches were further limited to the modification of protein termini or in-frame loop regions. In addition to its utility as a labeling handle, 3-formyl-phenylalanine has also been a useful approach to obtain cyclic peptides, via thiazolidine linkages that are stable in neutral pH. It is expected that the rich chemistry of aldehydes will result in the utilization of 3-formyl-phenylalanine for a number of different research efforts, some of which will be pursued in the Liu laboratory. Of the potential projects involving 3-formyl-phenylalanine, the most pressing endeavor is the successful selection of thiazolidine-based cyclic peptide inhibitors using

phage display technology, which could greatly facilitate drug discovery efforts upon optimization of the methodology. Finally, phenylalanine derivatives have played an indirect role in discovering key features of an important bioorthogonal reaction. In an attempt to optimize bioconjugation with azidophenylalanine in a phage display context, it was discovered that currently accepted models of CuAAC do not account for all kinetic data presented in this Dissertation. Based on this data, it is now proposed that CuAAC requires two distinct copper(I)-chelating events, with one stoichiometric equivalent of copper(I) chelating to azide, and another equivalent interacting with alkyne. Cycloaddition occurs after both copper-chelating events occur. It is expected that the kinetic data presented in Chapter V will be useful to researchers who wish to develop copper-chelating reagents with optimized properties to further enhance such an important bioorthogonal reaction. To conclude, the work presented in this Dissertation has expanded the techniques available to researchers who require such methodologies for the isolation, modification, and manipulation of chemically interesting proteins.

REFERENCES

(1) Wang, L.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2005**, *44*, 34.

(2) Schmeing, T. M.; Ramakrishnan, V. *Nature* **2009**, *461*, 1234.

(3) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. *Angew. Chem., Int. Ed.* **2005**, *44*, 7342.

(4) Johnson, L. N.; Lewis, R. J. *Chem. Rev.* **2001**, *101*, 2209.

(5) Haines, N.; Irvine, K. D. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 786.

(6) Okajima, T.; Xu, A.; Irvine, K. D. *J. Biol. Chem.* **2003**, *278*, 42340.

(7) Wells, L.; Whalen, S. A.; Hart, G. W. *Biochem. Biophys. Res. Commun.* **2003**, *302*, 435.

(8) Schnell, J. D.; Hicke, L. *J. Biol. Chem.* **2003**, *278*, 35857.

(9) Hershko, A.; Ciechanover, A. *Annu. Rev. Biochem.* **1998**, *67*, 425.

(10) Glazer, A. N. *Annu. Rev. Biochem.* **1970**, *39*, 101.

(11) Francis, M. B.; Wiley-VCH Verlag GmbH & Co. KGaA: 2007; Vol. 2, p 593.

(12) Tilley, S. D.; Francis, M. B. *J. Am. Chem. Soc.* **2006**, *128*, 1080.

(13) Kalia, J.; Raines, R. T. *Curr. Org. Chem.* **2010**, *14*, 138.

(14) Sletten, E. M.; Bertozzi, C. R. *Angew. Chem., Int. Ed.* **2009**, *48*, 6974.

(15) Byeon, J. Y.; Limpoco, F. T.; Bailey, R. C. *Langmuir* **2010**, *26*, 15430.

(16) Crisalli, P.; Hernandez, A. R.; Kool, E. T. *Bioconjugate Chem.* **2012**, *23*, 1969.

(17) Raindlova, V.; Pohl, R.; Sanda, M.; Hocek, M. *Angew. Chem., Int. Ed.* **2010**, *49*, 1064.

(18) Dixon, H. B. *Biochem. J.* **1962**, *83*, 91.

(19) Dixon, H. B.; Weitkamp, L. R. *Biochem. J.* **1962**, *84*, 462.

(20) Geschwind, I. I.; Hao Li, C. *Biochim. Biophys. Acta* **1954**, *15*, 442.

(21) McManus, J. F. *Nature* **1956**, *178*, 914.

(22) Kalia, J.; Raines, R. T. *Angew. Chem., Int. Ed.* **2008**, *47*, 7523.

(23) Agarwal, P.; van der Weijden, J.; Sletten, E. M.; Rabuka, D.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 46.

(24) Griffin, R. J. *Prog. Med. Chem.* **1994**, *31*, 121.

(25) Hendricks, S. B.; Pauling, L. *J. Am. Chem. Soc.* **1925**, *47*, 2904.

(26) Sidgwick, N. V.; Sutton, L. E.; Thomas, W. *J. Chem. Soc.* **1933**, 406.

(27) Saxon, E.; Bertozzi, C. R. *Science* **2000**, *287*, 2007.

(28) Soellner, M. B.; Nilsson, B. L.; Raines, R. T. *J. Am. Chem. Soc.* **2006**, *128*, 8820.

(29) Nilsson, B. L.; Kiessling, L. L.; Raines, R. T. *Org. Lett.* **2000**, *2*, 1939.

(30) Nilsson, B. L.; Kiessling, L. L.; Raines, R. T. *Org. Lett.* **2001**, *3*, 9.

(31) Soellner, M. B.; Tam, A.; Raines, R. T. *J. Org. Chem.* **2006**, *71*, 9824.

(32) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2001**, *40*, 2004.

(33) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2002**, *41*, 2596.

(34) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. *Angew. Chem., Int. Ed.* **2009**, *48*, 9879.

(35) Iha, R. K.; Wooley, K. L.; Nyström, A. M.; Burke, D. J.; Kade, M. J.; Hawker, C. J. *Chem. Rev.* **2009**, *109*, 5620.

(36) Lewis, W. G.; Green, L. G.; Grynszpan, F.; Radić, Z.; Carlier, P. R.; Taylor, P.; Finn, M. G.; Sharpless, K. B. *Angew. Chem.* **2002**, *114*, 1095.

(37) P. He, X.; Xie, J.; Tang, Y.; Li, J.; R. Chen, G. *Curr. Med. Chem.* **2012**, *19*, 2399.

(38) Palacin, T.; Khanh, H. L.; Jousselme, B.; Jegou, P.; Filoramo, A.; Ehli, C.; Guldi, D. M.; Campidelli, S. *J. Am. Chem. Soc.* **2009**, *131*, 15394.

(39) Uttamapinant, C.; Tangpeerachaikul, A.; Grecian, S.; Clarke, S.; Singh, U.; Slade, P.; Gee, K. R.; Ting, A. Y. *Angew. Chem., Int. Ed.* **2012**, *51*, 5852.

(40) Bevilacqua, V.; King, M.; Chaumontet, M.; Nothisen, M.; Gabillet, S.; Buisson, D.; Puente, C.; Wagner, A.; Taran, F. *Angew. Chem., Int. Ed.* **2014**, *53*, 5872.

(41) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2004**, *126*, 15046.

(42) Blackman, M. L.; Royzen, M.; Fox, J. M. *J. Am. Chem. Soc.* **2008**, *130*, 13518.

(43) Devaraj, N. K.; Weissleder, R.; Hilderbrand, S. A. *Bioconjugate Chem.* **2008**, *19*, 2297.

(44) Yang, J.; Seckute, J.; Cole, C. M.; Devaraj, N. K. *Angew. Chem., Int. Ed.* **2012**, *51*, 7476.

(45) Patterson, D. M.; Nazarova, L. A.; Xie, B.; Kamber, D. N.; Prescher, J. A. *J. Am. Chem. Soc.* **2012**, *134*, 18638.

(46) Lang, K.; Davis, L.; Wallace, S.; Mahesh, M.; Cox, D. J.; Blackman, M. L.; Fox, J. M.; Chin, J. W. *J. Am. Chem. Soc.* **2012**, *134*, 10317.

(47) Merrifield, R. B. In *Adv. Enzymol. Relat. Areas Mol. Biol.*; John Wiley & Sons, Inc.: 2006, p 221.

(48) Dawson, P. E.; Kent, S. B. H. *Annu. Rev. Biochem.* **2000**, *69*, 923.

(49) Muir, T. W.; Sondhi, D.; Cole, P. A. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 6705.

(50) Crick, F. H. *Symp. Soc. Exp. Biol.* **1958**, *12*, 138.

(51) Chapeville, F.; Lipmann, F.; Von Ehrenstein, G.; Weisblum, B.; Ray, W. J., Jr.; Benzer, S. *Proc. Natl. Acad. Sci. U. S. A.* **1962**, *48*, 1086.

(52) Johnson, A. E.; Woodward, W. R.; Herbert, E.; Menninger, J. R. *Biochemistry* **1976**, *15*, 569.

(53) Heckler, T. G.; Zama, Y.; Naka, T.; Hecht, S. M. *J. Biol. Chem.* **1983**, *258*, 4492.

(54) Heckler, T. G.; Chang, L. H.; Zama, Y.; Naka, T.; Chorghade, M. S.; Hecht, S. M. *Biochemistry* **1984**, *23*, 1468.

(55) Roesser, J. R.; Chorghade, M. S.; Hecht, S. M. *Biochemistry* **1986**, *25*, 6361.

(56) Baldini, G.; Martoglio, B.; Schachenmann, A.; Zugliani, C.; Brunner, J. *Biochemistry* **1988**, *27*, 7951.

(57) Heckler, T. G.; Roesser, J. R.; Xu, C.; Chang, P. I.; Hecht, S. M. *Biochemistry* **1988**, *27*, 7254.

(58) Cohen, G. N.; Munier, R. *Biochim. Biophys. Acta* **1956**, *21*, 592.

(59) Cohen, G. N.; Cowie, D. B. *C. R. Acad. Sci.* **1957**, *244*, 680.

(60) Cowie, D. B.; Cohen, G. N. *Biochim. Biophys. Acta* **1957**, *26*, 252.

(61) Hortin, G.; Boime, I. *Methods Enzymol.* **1983**, *96*, 777.

(62) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 19.

(63) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498.

(64) Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. *J. Am. Chem. Soc.* **2000**, *122*, 5010.

(65) Schultz, K. C.; Supekova, L.; Ryu, Y.; Xie, J.; Perera, R.; Schultz, P. G. *J. Am. Chem. Soc.* **2006**, *128*, 13984.

(66) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11020.

(67) Chin, J. W.; Schultz, P. G. *ChemBioChem* **2002**, *3*, 1135.

(68) Deiters, A.; Groff, D.; Ryu, Y.; Xie, J.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2006**, *45*, 2728.

(69) Stadtman, T. C. *Annu. Rev. Biochem.* **1996**, *65*, 83.

(70) Hao, B.; Gong, W.; Ferguson, T. K.; James, C. M.; Krzycki, J. A.; Chan, M. K. *Science* **2002**, *296*, 1462.

(71) Srinivasan, G.; James, C. M.; Krzycki, J. A. *Science (New York, N.Y.)* **2002**, *296*, 1459.

(72) Zhang, Y.; Baranov, P. V.; Atkins, J. F.; Gladyshev, V. N. *J. Biol. Chem.* **2005**, *280*, 20740.

(73) Polycarpo, C. R.; Herring, S.; Berube, A.; Wood, J. L.; Soll, D.; Ambrogelly, A. *FEBS Lett.* **2006**, *580*, 6695.

(74) Ou, W.; Uno, T.; Chiu, H. P.; Grunewald, J.; Cellitti, S. E.; Crossgrove, T.; Hao, X.; Fan, Q.; Quinn, L. L.; Patterson, P.; Okach, L.; Jones, D. H.; Lesley, S. A.; Brock, A.; Geierstanger, B. H. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 10437.

(75) Lang, K.; Davis, L.; Torres-Kolbus, J.; Chou, C.; Deiters, A.; Chin, J. W. *Nature Chem.* **2012**, *4*, 298.

(76) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. *Chem. Biol.* **2008**, *15*, 1187.

(77) Nguyen, D. P.; Elliott, T.; Holt, M.; Muir, T. W.; Chin, J. W. *J. Am. Chem. Soc.* **2011**, *133*, 11418.

(78) Lee, Y. J.; Wu, B.; Raymond, J. E.; Zeng, Y.; Fang, X.; Wooley, K. L.; Liu, W. R. *ACS Chem. Biol.* **2013**, *8*, 1664.

(79) Darko, A.; Wallace, S.; Dmitrenko, O.; Machovina, M. M.; Mehl, R. A.; Chin, J. W.; Fox, J. M. *Chem. Sci.* **2014**, *5*, 3770.

(80) Kurra, Y.; Odoi, K. A.; Lee, Y. J.; Yang, Y.; Lu, T.; Wheeler, S. E.; Torres-Kolbus, J.; Deiters, A.; Liu, W. R. *Bioconjugate Chem.* **2014**, *25*, 1730.

(81) Liu, W. R.; Wang, Y.-S.; Wan, W. *Mol. Biosyst.* **2011**, *7*, 38.

(82) Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. *Chem. Sci.* **2015**, *6*, 50.

(83) Wang, Y. S.; Russell, W. K.; Wang, Z.; Wan, W.; Dodd, L. E.; Pai, P. J.; Russell, D. H.; Liu, W. R. *Mol. Biosyst.* **2011**, *7*, 714.

(84) Ko, J.-h.; Wang, Y.-S.; Nakamura, A.; Guo, L.-T.; Söll, D.; Umehara, T. *FEBS Lett.* **2013**, *587*, 3243.

(85) Wang, Y.-S.; Russell, W. K.; Wang, Z.; Wan, W.; Dodd, L. E.; Pai, P.-J.; Russell, D. H.; Liu, W. R. *Mol. BioSyst.* **2011**, *7*, 714.

(86) Gautier, A.; Deiters, A.; Chin, J. W. *J. Am. Chem. Soc.* **2011**, *133*, 2124.

(87) Zhang, M.; Lin, S.-X.; Song, X.-W.; Liu, J.; Fu, Y.; Ge, X.; Fu, X.-M.; Chang, Z.-Y.; Chen, P.-R. *Nat. Chem. Biol.* **2011**, *7*, 671.

(88) Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D. T.; Goddard, W. A., III; DeGrado, W. F.; Tirrell, D. A. *Biochemistry* **2001**, *40*, 2790.

(89) Liu, C. C.; Schultz, P. G. *Annu. Rev. Biochem.* **2010**, *79*, 413.

(90) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498.

(91) Xie, J.; Schultz, P. G. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 775.

(92) Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124*, 9026.

(93) Srinivasan, G.; James, C. M.; Krzycki, J. A. *Science* **2002**, *296*, 1459.

(94) Blight, S. K.; Larue, R. C.; Mahapatra, A.; Longstaff, D. G.; Chang, E.; Zhao, G.; Kang, P. T.; Green-Church, K. B.; Chan, M. K.; Krzycki, J. A. *Nature* **2004**, *431*, 333.

(95) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat. Chem. Biol.* **2008**, *4*, 232.

(96) Wan, W.; Huang, Y.; Wang, Z.; Russell, W. K.; Pai, P.-J.; Russell, D. H.; Liu, W. R. *Angew. Chem., Int. Ed.* **2010**, *49*, 3211.

(97) Greiss, S.; Chin, J. W. *J. Am. Chem. Soc.* **2011**, *133*, 14196.

(98) Hancock, S. M.; Uprety, R.; Deiters, A.; Chin, J. W. *J. Am. Chem. Soc.* **2010**, *132*, 14819.

(99) Mukai, T.; Kobayashi, T.; Hino, N.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S. *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818.

(100) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. *Chem. Biol.* **2008**, *15*, 1187.

(101) Parrish, A. R.; She, X.; Xiang, Z.; Coin, I.; Shen, Z.; Briggs, S. P.; Dillin, A.; Wang, L. *ACS Chem. Biol.* **2012**, *7*, 1292.

(102) Chen, P. R.; Groff, D.; Guo, J.; Ou, W.; Cellitti, S.; Geierstanger, B. H.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2009**, *48*, 4052.

(103) Chou, C.; Uprety, R.; Davis, L.; Chin, J. W.; Deiters, A. *Chem. Sci.* **2011**, *2*, 480.

(104) Wang, Y.-S.; Wu, B.; Wang, Z.-Y.; Huang, Y.; Wan, W.; Russell, W. K.; Pai, P.-J.; Moe, Y. N.; Russell, D. H.; Liu, W.-S. R. *Mol. BioSyst.* **2010**, *6*, 1557.

(105) Lee, Y.-J.; Wu, B.; Raymond, J. E.; Zeng, Y.; Fang, X.; Wooley, K. L.; Liu, W. R. *ACS Chem. Biol.* **2013**, *8*, 1664.

(106) Fekner, T.; Li, X.; Lee, M. M.; Chan, M. K. *Angew. Chem., Int. Ed.* **2009**, *48*, 1633.

(107) Li, X.; Fekner, T.; Ottesen, J. J.; Chan, M. K. *Angew. Chem., Int. Ed.* **2009**, *48*, 9184.

(108) Umehara, T.; Kim, J.; Lee, S.; Guo, L.-T.; Soll, D.; Park, H.-S. *FEBS Lett.* **2012**, *586*, 729.

(109) Polycarpo, C. R.; Herring, S.; Berube, A.; Wood, J. L.; Soell, D.; Ambrogelly, A. *FEBS Lett.* **2006**, *580*, 6695.

(110) Plass, T.; Milles, S.; Koehler, C.; Schultz, C.; Lemke, E. A. *Angew. Chem., Int. Ed.* **2011**, *50*, 3878.

(111) Nguyen, D. P.; Lusic, H.; Neumann, H.; Kapadnis, P. B.; Deiters, A.; Chin, J. W. *J. Am. Chem. Soc.* **2009**, *131*, 8720.

(112) Wang, Y.-S.; Fang, X.; Wallace, A. L.; Wu, B.; Liu, W. R. *J. Am. Chem. Soc.* **2012**, *134*, 2950.

(113) Wang, Y.-S.; Fang, X.; Chen, H.-Y.; Wu, B.; Wang, Z. U.; Hilty, C.; Liu, W. R. *ACS Chem. Biol.* **2013**, *8*, 405.

(114) Jewett, J. C.; Sletten, E. M.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2010**, *132*, 3688.

(115) Wang, Y. S.; Fang, X.; Wallace, A. L.; Wu, B.; Liu, W. R. *J. Am. Chem. Soc.* **2012**, *134*, 2950.

(116) Wang, Y. S.; Fang, X.; Chen, H. Y.; Wu, B.; Wang, Z. U.; Hilty, C.; Liu, W. R. *ACS Chem. Biol.* **2013**, *8*, 405.

(117) Wu, B.; Wang, Z.; Huang, Y.; Liu, W. R. *ChemBioChem* **2012**, *13*, 1405.

(118) Cabiddu, M. G.; Cadoni, E.; De Montis, S.; Fattuoni, C.; Melis, S.; Usai, M. *Tetrahedron* **2003**, *59*, 4383.

(119) Archer, A. W.; Claret, P. A.; Hayman, D. F. *J. Chem. Soc. B.* **1971**, 1231.

(120) Drew, S. L.; Lawrence, A. L.; Sherburn, M. S. *Angew. Chem., Int. Ed.* **2013**, *52*, 4221.

(121) Humphrey, C. E.; Furegati, M.; Laumen, K.; La Vecchia, L.; Leutert, T.; Müller-Hartwieg, J. C. D.; Vögtle, M. *Org. Process Res. Dev.* **2007**, *11*, 1069.

(122) Carroll, F. I.; Blough, B. E.; Abraham, P.; Mills, A. C.; Holleman, J. A.; Wolckenhauer, S. A.; Decker, A. M.; Landavazo, A.; McElroy, K. T.; Navarro, H. A.; Gatch, M. B.; Forster, M. J. *J. Med. Chem.* **2009**, *52*, 6768.

(123) Liu, Y.; Yao, B.; Deng, C. L.; Tang, R. Y.; Zhang, X. G.; Li, J. H. *Org. Lett.* **2011**, *13*, 2184.

(124) Meltzer, P. C.; Butler, D.; Deschamps, J. R.; Madras, B. K. *J. Med. Chem.* **2006**, *49*, 1420.

(125) Ruan, J.; Saidi, O.; Iggo, J. A.; Xiao, J. *J. Am. Chem. Soc.* **2008**, *130*, 10510.

(126) McFarland, J. M.; Francis, M. B. *J. Am. Chem. Soc.* **2005**, *127*, 13490.

(127) McFarland, J. M.; Joshi, N. S.; Francis, M. B. *J. Am. Chem. Soc.* **2008**, *130*, 7639.

(128) Han, M. J.; Xiong, D. C.; Ye, X. S. *Chem. Commun.* **2012**, *48*, 11079.

(129) Agarwal, P.; Kudirka, R.; Albers, A. E.; Barfield, R. M.; de Hart, G. W.; Drake, P. M.; Jones, L. C.; Rabuka, D. *Bioconjugate Chem.* **2013**, *24*, 846.

(130) Witus, L. S.; Netirojjanakul, C.; Palla, K. S.; Muehl, E. M.; Weng, C. H.; Iavarone, A. T.; Francis, M. B. *J. Am. Chem. Soc.* **2013**, *135*, 17223.

(131) Carrico, I. S.; Carlson, B. L.; Bertozzi, C. R. *Nat. Chem. Biol.* **2007**, *3*, 321.

(132) Rush, J. S.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2008**, *130*, 12240.

(133) Wu, P.; Shui, W.; Carlson, B. L.; Hu, N.; Rabuka, D.; Lee, J.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 3000.

(134) Hudak, J. E.; Yu, H. H.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2011**, *133*, 16127.

(135) El-Mahdi, O.; Melnyk, O. *Bioconjugate Chem.* **2013**, *24*, 735.

(136) Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 56.

(137) Zhang, Z.; Smith, B. A.; Wang, L.; Brock, A.; Cho, C.; Schultz, P. G. *Biochemistry* **2003**, *42*, 6735.

(138) Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Wang, Z.; Russell, D. H.; Liu, W. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 878.

(139) Wu, B.; Wang, Z.; Huang, Y.; Liu, W. R. *ChemBioChem* **2012**, *13*, 1405.

(140) Zeng, H.; Xie, J.; Schultz, P. G. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5356.

(141) Dirksen, A.; Hackeng, T. M.; Dawson, P. E. *Angew. Chem., Int. Ed.* **2006**, *45*, 7581.

(142) Dirksen, A.; Dawson, P. E. *Bioconjugate Chem.* **2008**, *19*, 2543.

(143) Zeng, Y.; Ramya, T. N.; Dirksen, A.; Dawson, P. E.; Paulson, J. C. *Nat. Methods* **2009**, *6*, 207.

(144) Wendeler, M.; Grinberg, L.; Wang, X.; Dawson, P. E.; Baca, M. *Bioconjugate Chem.* **2013**.

(145) Tharp, J. M.; Wang, Y. S.; Lee, Y. J.; Yang, Y.; Liu, W. R. *ACS Chem. Biol.* **2014**.

(146) Tuley, A.; Wang, Y. S.; Fang, X.; Kurra, Y.; Rezenom, Y. H.; Liu, W. R. *Chem. Commun.* **2014**, *50*, 2673.

(147) Reddington, S. C.; Tippmann, E. M.; Jones, D. D. *Chem. Commun.* **2012**, *48*, 8419.

(148) Pedelacq, J. D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. *Nat. Biotechnol.* **2006**, *24*, 79.

(149) Wang, X. S.; Lee, Y.-J.; Liu, W. R. *Chem. Commun.* **2014**, *50*, 3176.

(150) Borrmann, A.; Milles, S.; Plass, T.; Dommerholt, J.; Verkade, J. M.; Wiessler, M.; Schultz, C.; van Hest, J. C.; van Delft, F. L.; Lemke, E. A. *ChemBioChem* **2012**, *13*, 2094.

(151) Plass, T.; Milles, S.; Koehler, C.; Szymanski, J.; Mueller, R.; Wiessler, M.; Schultz, C.; Lemke, E. A. *Angew. Chem., Int. Ed.* **2012**, *51*, 4166.

(152) Lorello, G. R.; Legault, M. C.; Rakic, B.; Bisgaard, K.; Pezacki, J. P. *Bioorg. Chem.* **2008**, *36*, 105.

(153) Ossipov, D. A.; Yang, X.; Varghese, O.; Kootala, S.; Hilborn, J. *Chem. Commun.* **2010**, *46*, 8368.

(154) Rotstein, B. H.; Rai, V.; Hili, R.; Yudin, A. K. *Nat. Protoc.* **2010**, *5*, 1813.

(155) Iyer, G.; Pinaud, F.; Xu, J.; Ebenstein, Y.; Li, J.; Chang, J.; Dahan, M.; Weiss, S. *Bioconjugate Chem.* **2011**, *22*, 1006.

(156) Cohen, J. D.; Zou, P.; Ting, A. Y. *ChemBioChem* **2012**, *13*, 888.

(157) Raindlova, V.; Pohl, R.; Hocek, M. *Chem. Eur. J.* **2012**, *18*, 4080.

(158) Dhal, P. K.; Polomoscanik, S. C.; Gianolio, D. A.; Starremans, P. G.; Busch, M.; Alving, K.; Chen, B.; Miller, R. J. *Bioconjugate Chem.* **2013**, *24*, 865.

(159) Ossipov, D.; Kootala, S.; Yi, Z.; Yang, X.; Hilborn, J. *Macromolecules* **2013**, *46*, 4105.

(160) Tanaka, K.; Nakamoto, Y.; Siwu, E. R.; Pradipta, A. R.; Morimoto, K.; Fujiwara, T.; Yoshida, S.; Hosoya, T.; Tamura, Y.; Hirai, G.; Sodeoka, M.; Fukase, K. *Org. Biomol. Chem.* **2013**, *11*, 7326.

(161) Wang, S.; Oommen, O. P.; Yan, H.; Varghese, O. P. *Biomacromolecules* **2013**, *14*, 2427.

(162) Bookser, B. C.; Bruice, T. C. *J. Am. Chem. Soc.* **1991**, *113*, 4208.

(163) Hammill, J. T.; Miyake-Stoner, S.; Hazen, J. L.; Jackson, J. C.; Mehl, R. A. *Nat. Protoc.* **2007**, *2*, 2601.

(164) Packer, M. S.; Liu, D. R. *Nat. Rev. Genet.* **2015**, *16*, 379.

(165) Smith, G. P.; Petrenko, V. A. *Chem. Rev.* **1997**, *97*, 391.

(166) Smith, G. P. *Science* **1985**, *228*, 1315.

(167) Tian, F.; Tsao, M. L.; Schultz, P. G. *J. Am. Chem. Soc.* **2004**, *126*, 15962.

(168) Pu, Y.; Martin, F. M.; Vederas, J. C. *J. Org. Chem.* **1991**, *56*, 1280.

(169) Lall, M. S.; Ramtohul, Y. K.; James, M. N.; Vederas, J. C. *J. Org. Chem.* **2002**, *67*, 1536.

(170) Horton, D. A.; Bourne, G. T.; Smythe, M. L. *J. Comput. Aided Mol. Des.* **2002**, *16*, 415.

(171) Rezai, T.; Yu, B.; Millhauser, G. L.; Jacobson, M. P.; Lokey, R. S. *J. Am. Chem. Soc.* **2006**, *128*, 2510.

(172) Gentilucci, L.; De Marco, R.; Cerisoli, L. *Curr. Pharm. Des.* **2010**, *16*, 3185.

(173) Heinis, C.; Rutherford, T.; Freund, S.; Winter, G. *Nat. Chem. Biol.* **2009**, *5*, 502.

(174) Chen, S.; Rentero Rebollo, I.; Buth, S. A.; Morales-Sanfrutos, J.; Touati, J.; Leiman, P. G.; Heinis, C. *J. Am. Chem. Soc.* **2013**, *135*, 6562.

(175) Saiz, C.; Wipf, P.; Manta, E.; Mahler, G. *Org. Lett.* **2009**, *11*, 3170.

(176) Zhang, L.; Tam, J. P. *Anal. Biochem.* **1996**, *233*, 87.

(177) Ren, H.; Xiao, F.; Zhan, K.; Kim, Y.-P.; Xie, H.; Xia, Z.; Rao, J. *Angew. Chem., Int. Ed.* **2009**, *48*, 9658.

(178) Wang, Z.; Gu, C.; Colby, T.; Shindo, T.; Balamurugan, R.; Waldmann, H.; Kaiser, M.; van der Hoorn, R. A. *Nat. Chem. Biol.* **2008**, *4*, 557.

(179) Solorzano, C.; Antonietti, F.; Duranti, A.; Tontini, A.; Rivara, S.; Lodola, A.; Vacondio, F.; Tarzia, G.; Piomelli, D.; Mor, M. *J. Med. Chem.* **2010**, *53*, 5770.

(180) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596.

(181) Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G. *J. Am. Chem. Soc.* **2003**, *125*, 3192.

(182) Uttamapinant, C.; Sanchez, M. I.; Liu, D. S.; Yao, J. Z.; Ting, A. Y. *Nat. Protoc.* **2013**, *8*, 1620.

(183) Wu, P.; Chen, X.; Hu, N.; Tam, U. C.; Blixt, O.; Zettl, A.; Bertozzi, C. R. *Angew. Chem., Int. Ed.* **2008**, *47*, 5022.

(184) Steinmetz, N. F.; Hong, V.; Spoerke, E. D.; Lu, P.; Breitenkamp, K.; Finn, M. G.; Manchester, M. *J. Am. Chem. Soc.* **2009**, *131*, 17093.

(185) de Miguel, G.; Wielopolski, M.; Schuster, D. I.; Fazio, M. A.; Lee, O. P.; Haley, C. K.; Ortiz, A. L.; Echegoyen, L.; Clark, T.; Guldi, D. M. *J. Am. Chem. Soc.* **2011**, *133*, 13036.

(186) Himo, F.; Lovell, T.; Hilgraf, R.; Rostovtsev, V. V.; Noodleman, L.; Sharpless, K. B.; Fokin, V. V. *J. Am. Chem. Soc.* **2004**, *127*, 210.

(187) Rodionov, V. O.; Fokin, V. V.; Finn, M. G. *Angew. Chem. Int. Ed.* **2005**, *44*, 2210.

(188) Nolte, C.; Mayer, P.; Straub, B. F. *Angew. Chem. Int. Ed.* **2007**, *46*, 2101.

(189) Rodionov, V. O.; Presolski, S. I.; Díaz Díaz, D.; Fokin, V. V.; Finn, M. G. *J. Am. Chem. Soc.* **2007**, *129*, 12705.

(190) Kuang, G. C.; Guha, P. M.; Brotherton, W. S.; Simmons, J. T.; Stankee, L. A.; Nguyen, B. T.; Clark, R. J.; Zhu, L. *J. Am. Chem. Soc.* **2011**, *133*, 13984.

(191) Worrell, B. T.; Malik, J. A.; Fokin, V. V. *Science* **2013**, *340*, 457.

(192) Iacobucci, C.; Reale, S.; Gal, J.-F.; De Angelis, F. *Angew. Chem. Int. Ed.* **2015**, *54*, 3065.

(193) Berg, R.; Straub, B. F. *Beilstein J. Org. Chem.* **2013**, *9*, 2715.

(194) Jin, L.; Tolentino, D. R.; Melaimi, M.; Bertrand, G. *Sci. Adv.* **2015**, *1*.

(195) Kuang, G. C.; Michaels, H. A.; Simmons, J. T.; Clark, R. J.; Zhu, L. *J. Org. Chem.* **2010**, *75*, 6540.

(196) Sivakumar, K.; Xie, F.; Cash, B. M.; Long, S.; Barnhill, H. N.; Wang, Q. *Org. Lett.* **2004**, *6*, 4603.

(197) Sivakumar, K.; Xie, F.; Cash, B. M.; Long, S.; Barnhill, H. N.; Wang, Q. *Org. Lett.* **2004**, *6*, 4603.

(198) Chtchigrovsky, M.; Primo, A.; Gonzalez, P.; Molvinger, K.; Robitzer, M.; Quignard, F.; Taran, F. *Angew. Chem., Int. Ed.* **2009**, *48*, 5916.

(199) Soriano Del Amo, D.; Wang, W.; Jiang, H.; Besanceney, C.; Yan, A. C.; Levy, M.; Liu, Y.; Marlow, F. L.; Wu, P. *J. Am. Chem. Soc.* **2010**, *132*, 16893.

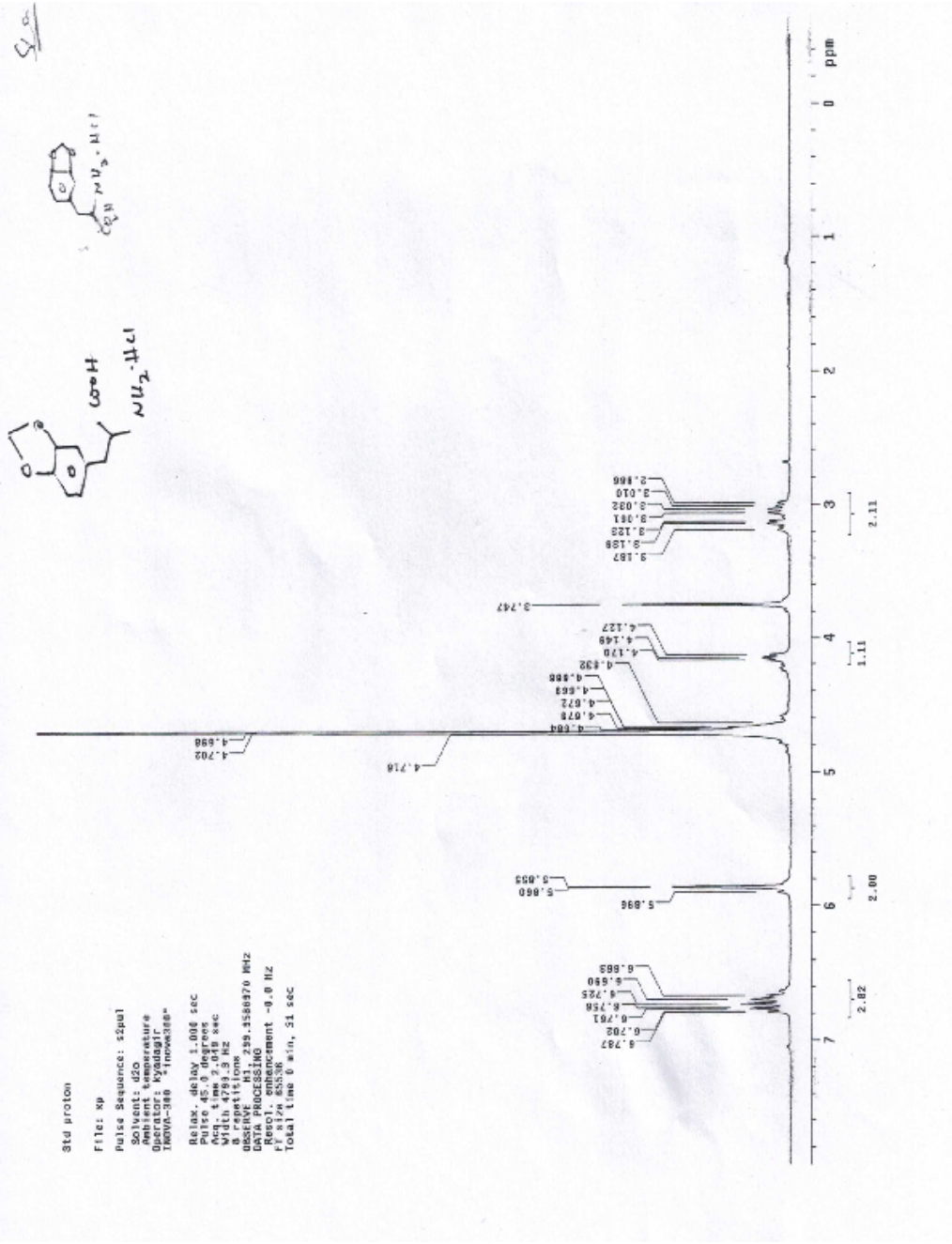
(200) Besanceney-Webler, C.; Jiang, H.; Zheng, T.; Feng, L.; Soriano del Amo, D.; Wang, W.; Klivansky, L. M.; Marlow, F. L.; Liu, Y.; Wu, P. *Angew. Chem., Int. Ed.* **2011**, *50*, 8051.

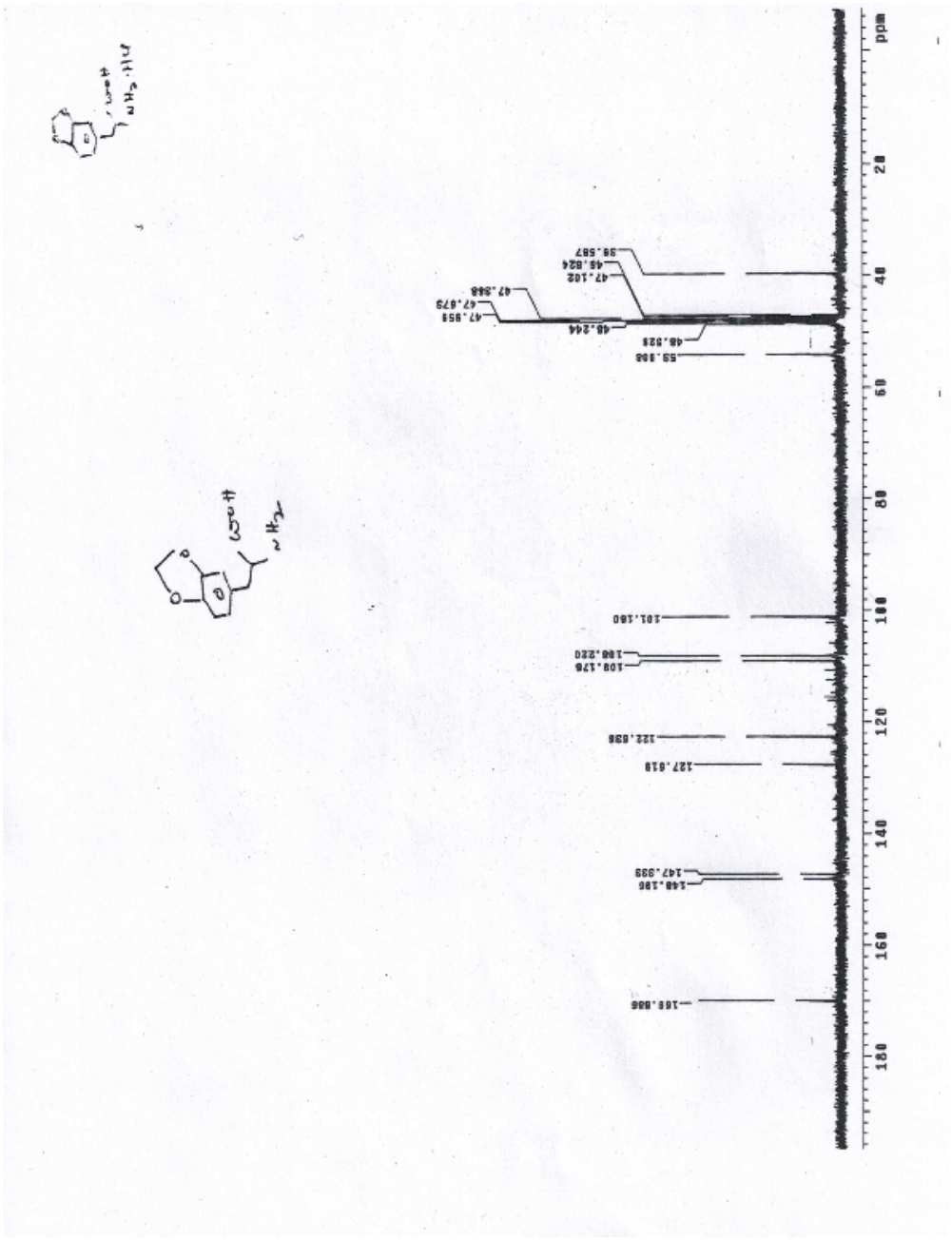
(201) Dong, J.; Xun, Z.; Zeng, Y.; Yu, T.; Han, Y.; Chen, J.; Li, Y. Y.; Yang, G.; Li, Y. *Chem. Eur. J.* **2013**, *19*, 7931.

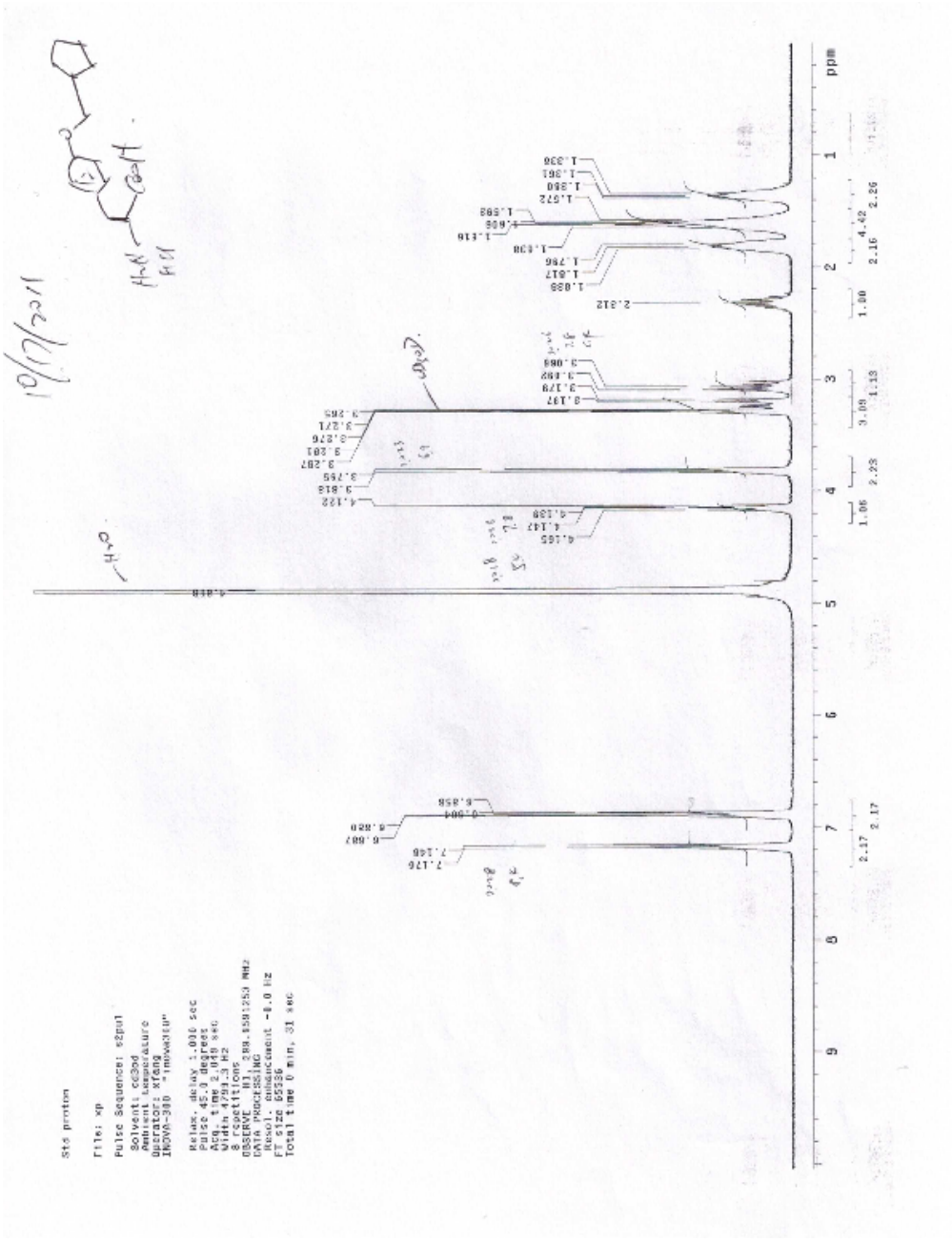
(202) Shi, F.; Waldo, J. P.; Chen, Y.; Larock, R. C. *Org. Lett.* **2008**, *10*, 2409.

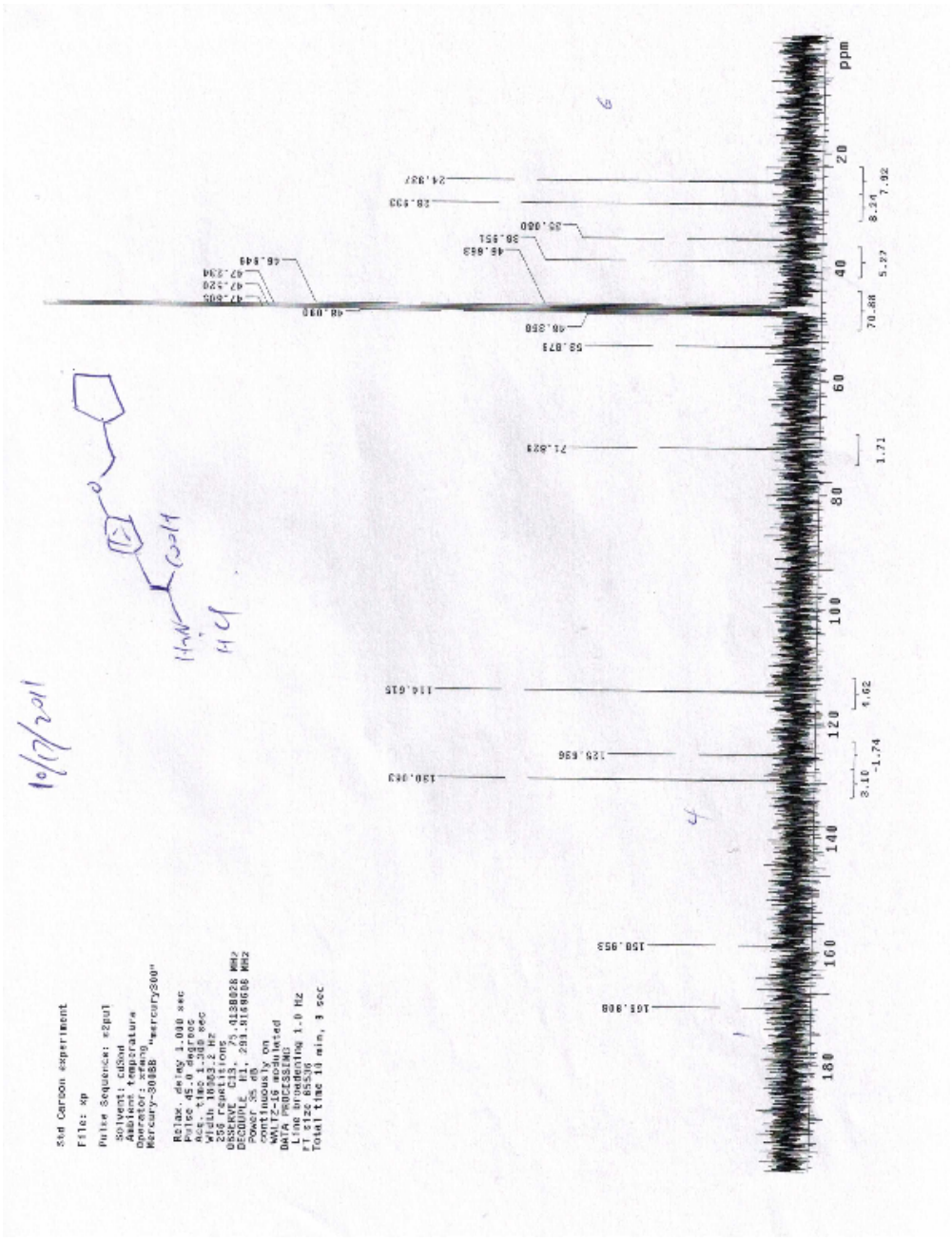
APPENDIX A

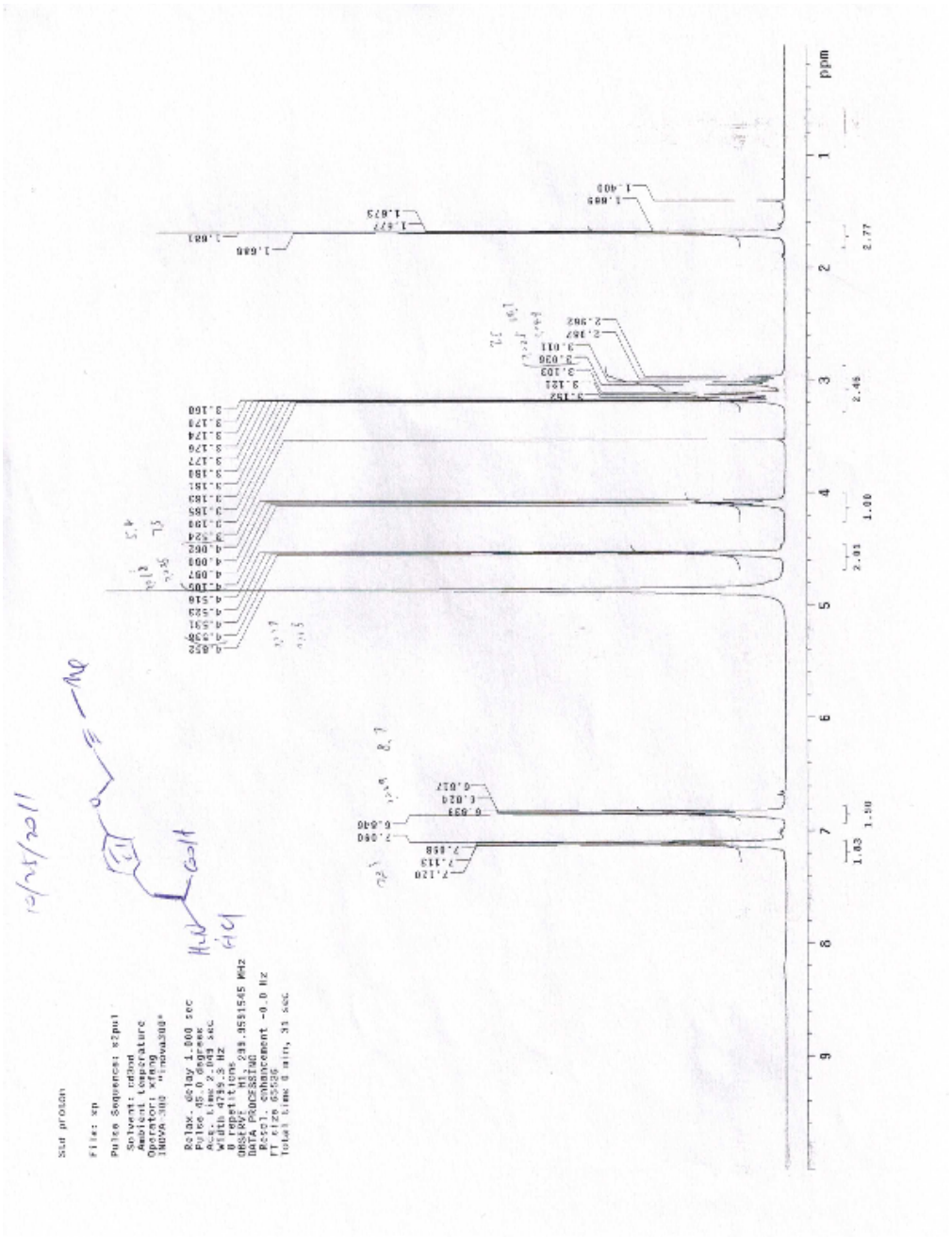
NMR DATA FOR SECTION 2

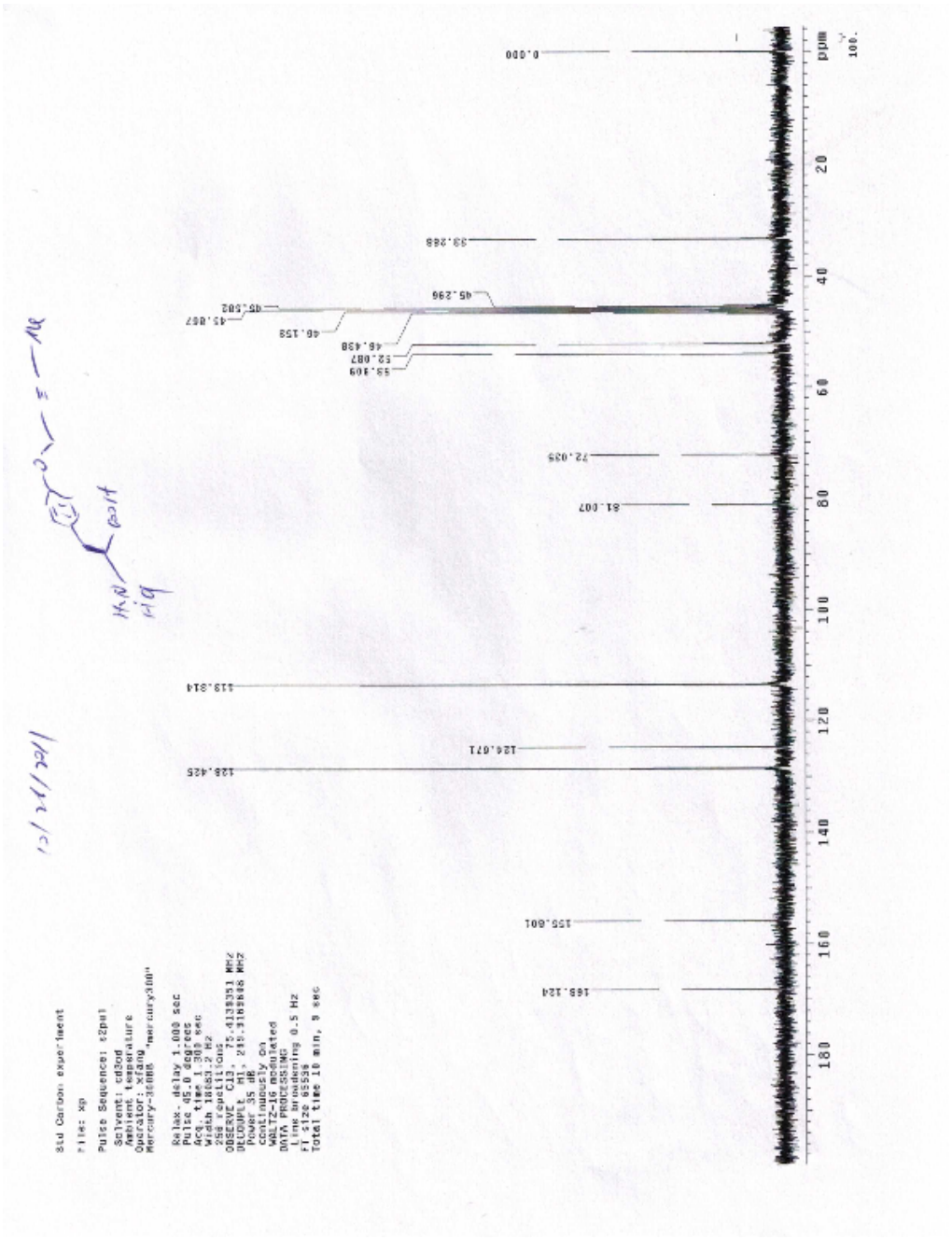


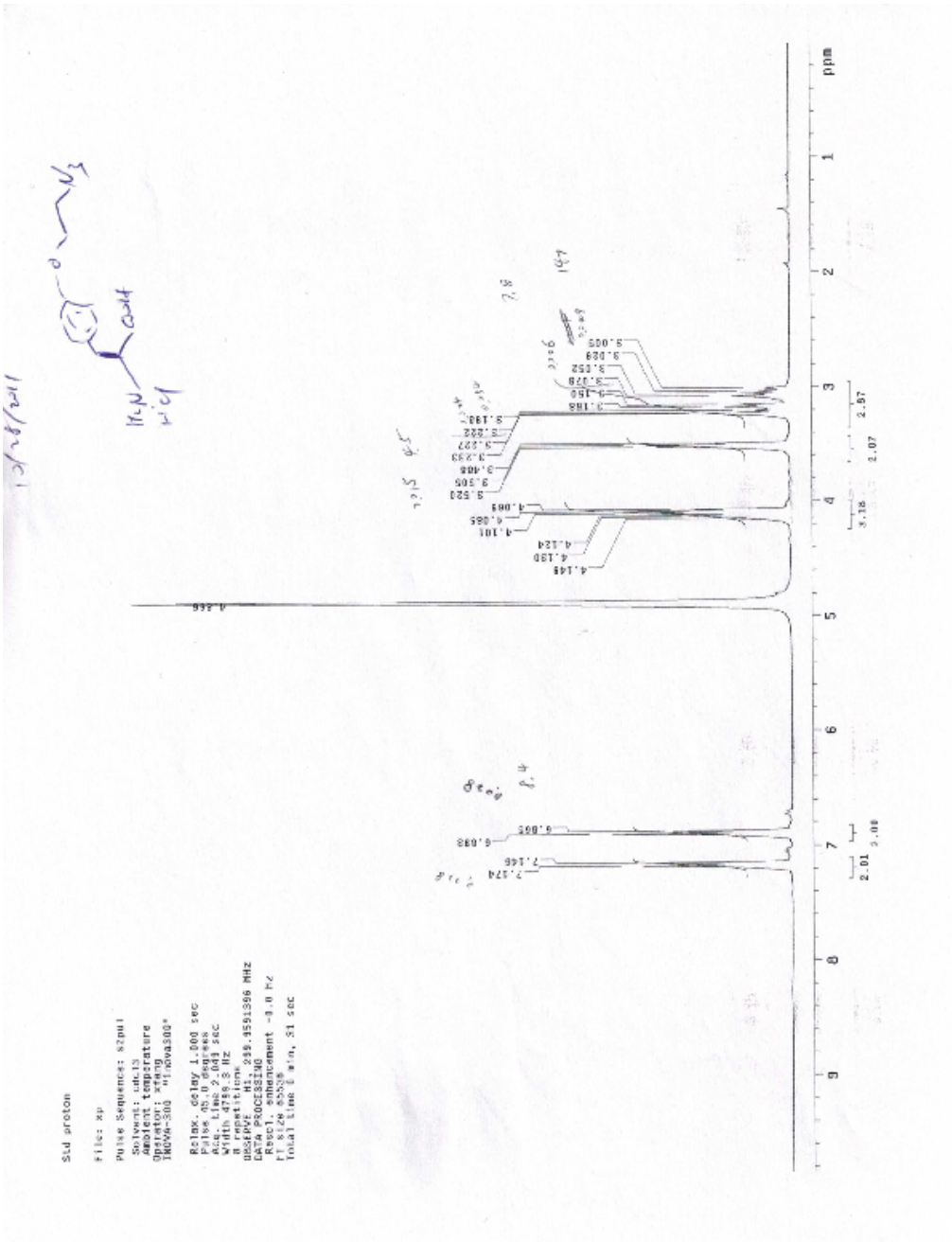

ppm
20
40
60
80
100
120
140
160
180
59.886
48.325
48.244
47.959
47.679
47.348
47.102
45.824
38.587
101.180
108.176
108.220
122.636
127.618
147.999
148.186
168.886

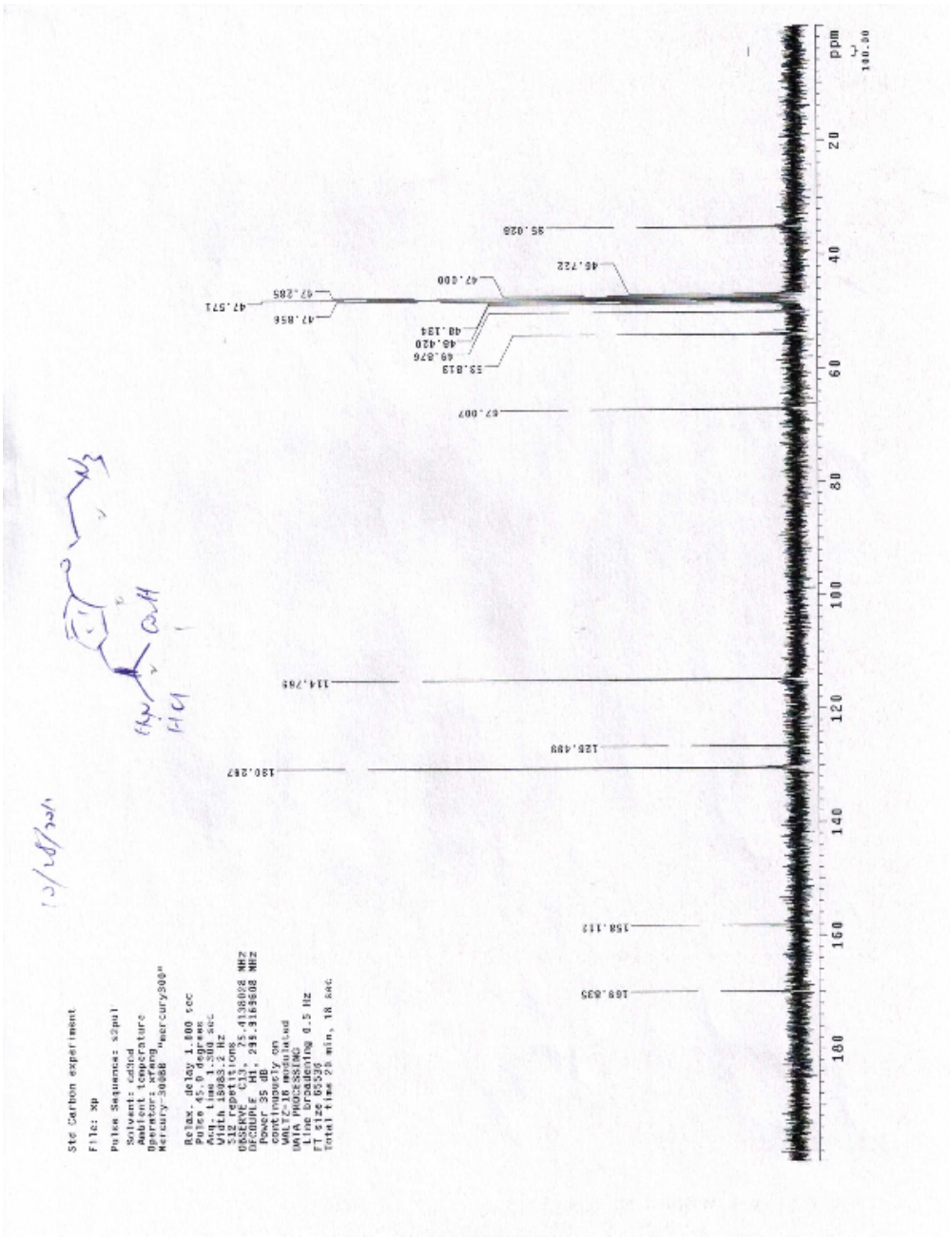

Sts proton
File: xp
Pulse Sequence: s2pul
Solvent: cd3od
Ambient temperature
Operator: xmoon
INOVA-300 "mnmoken"
Relax. delay 1.000 sec
Pulse 45.0 degrees
Acq. time 2.049 sec
Width 4811.1 Hz
8 repetitions
OBSERVE H1, 299.8581253 MHz
DATA PROCESSING
Resol. enhancement -0.6 Hz
FT size 65536
Total time 0 min, 31 sec
4.818
3.312
3.287
3.281
3.276
3.271
3.265
3.818
3.795
4.122
4.165
4.147
4.128
7.176
7.156
6.887
6.820
6.858
6.864
3.137
3.179
3.092
3.089
2.812
1.088
1.817
1.786
1.638
1.618
1.606
1.588
1.572
1.360
1.361
1.336
ppm

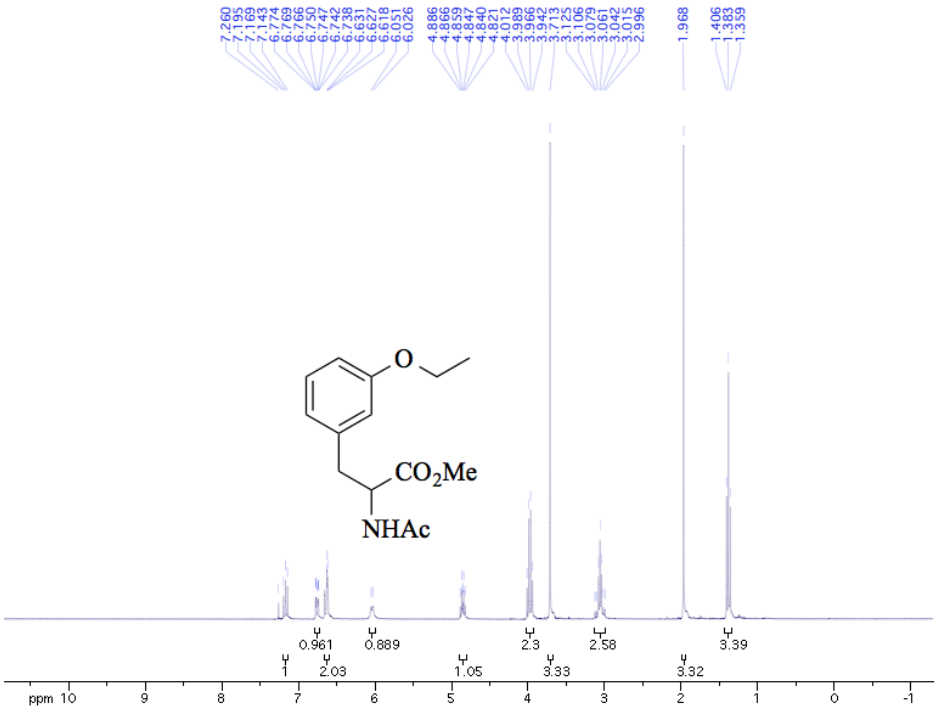

Std Carbon experiment
File: xp
Pulse Sequence: s2pul
Solvent: cdcl3nd
Ambient temperature
Operator: ****
Mercury-300BB "mercury300"
Relax. delay 1.000 sec
Pulse 45.0 degree
Acq. time 1.300 sec
Width 18153.2 Hz
256 repetitions
OBSERVE C13, 75.4138028 MHz
DECOUPLE H1, 299.9169698 MHz
POWER 35 dB
continuously ON
WALTZ-16 modulated
DATA PROCESSING
Line broadening 1.0 Hz
FT size 65536
Total time 10 min, 3 sec
10/17/2011
ppm
24.827
28.503
35.080
38.951
48.853
48.080
48.350
58.873
66.949
47.808
47.520
47.234
71.829
114.615
125.696
130.063
150.955
165.808
70.88
5.27
8.34
7.92
1.71
4.62
3.10
-1.74

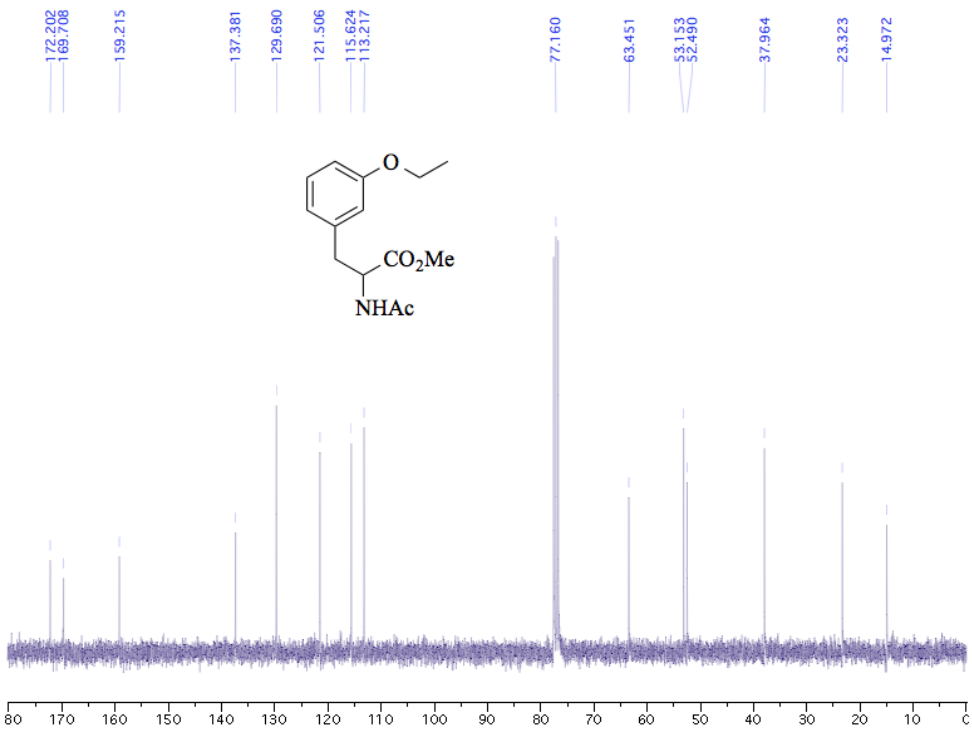

Std proton
File: xp
Pulse Sequence: s2pul
Solvent: cdcl3
Ambient temperature
Operator: xfeng
INOVA-300 "mnova300"
Relax. delay 1.800 sec
Pulse 45.0 degrees
Acq. time 2.043 sec
Width 4798.2 Hz
8 repetitions
OBSERVE H1, 299.9551545 MHz
DATA PROCESSING
Resol. enhancement -0.0 Hz
FT size 65536
Total time 0 min, 33 sec
ppm

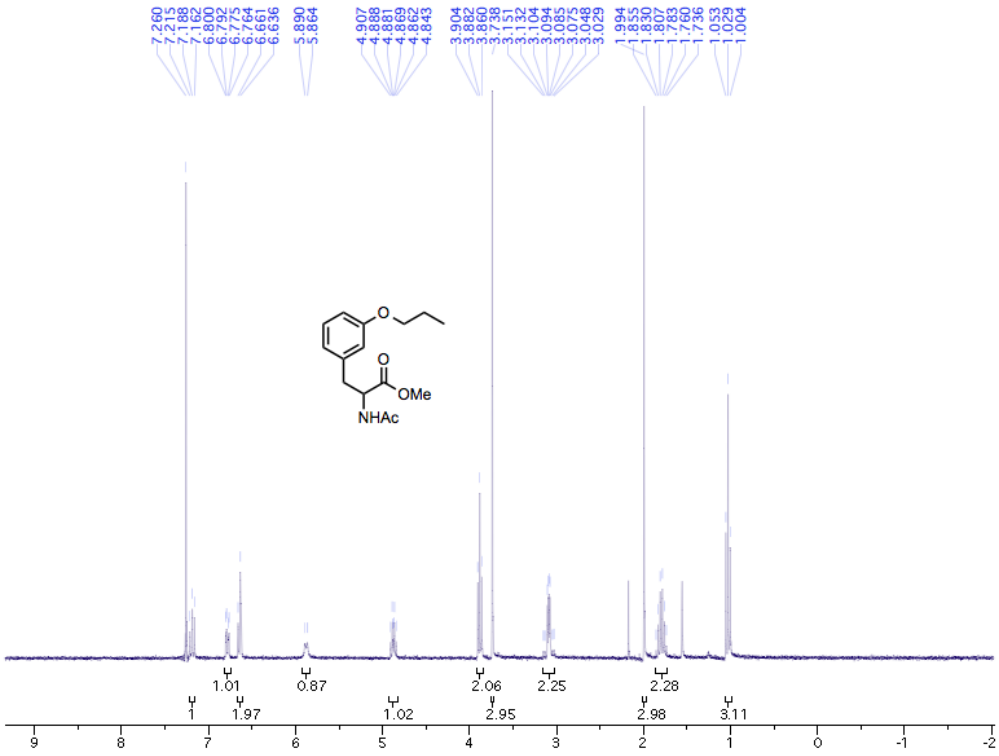

Std Carbon experiment
File: xp
Pulse Sequence: s2pul
Solvent: cd3od
Ambient temperature
Operator: xxxxxx
Mercury-300BB "mercury300"
Relax. delay 1.000 sec
Pulse 45.0 degrees
Acq. time 1.300 sec
Width 18833.2 Hz
2 repetitions
OBSERVE C13, 75.4138351 MHz
DECOUPL H1, 285.9188848 MHz
Power 35 dB
Continuously On
WALTZ16 modulated
DATA PROCESSING
Line broadening 0.5 Hz
FT size 65536
Total time 10 min, 8 sec
0.000
33.288
45.296
45.502
45.867
46.158
46.458
52.087
53.909
72.035
81.007
119.814
126.671
128.425
155.601
169.124
ppm
100.
20
40
60
80
100
120
140
160
180

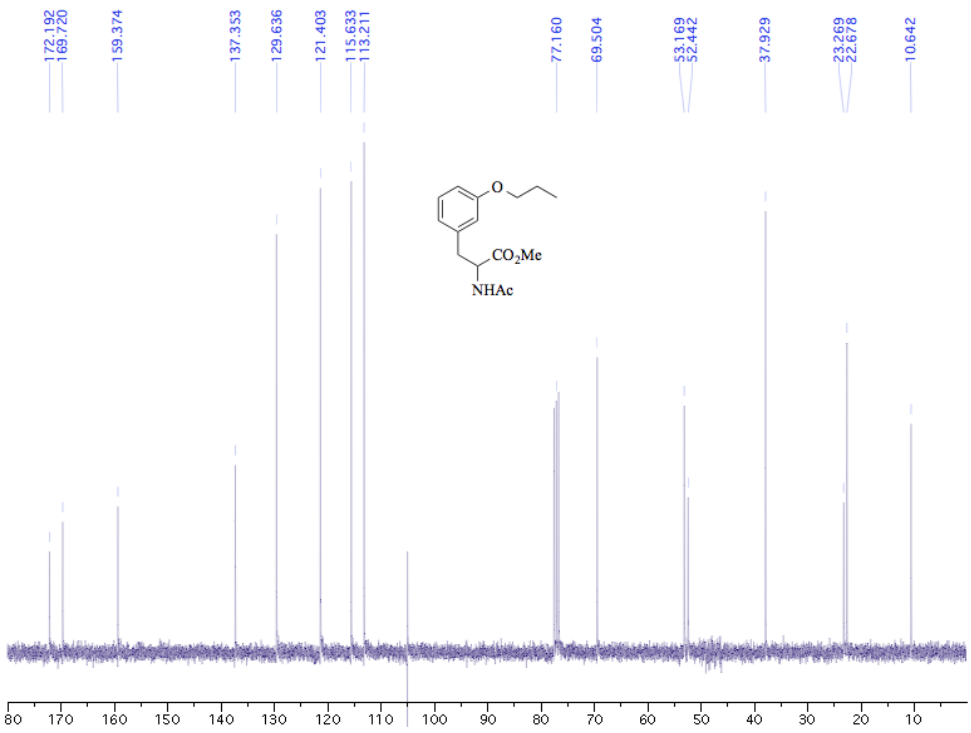

Std proton
File: xp
Pulse Sequence: s2pul
Solvent: cdcl3
Ambient temperature
Operator: xnmg
INOVA-500 "mercury500"
Relax. delay 1.000 sec
Pulse 45.0 degrees
Acq. time 2.049 sec
Width 4789.8 Hz
1 repetitions
OBSERVE H1, 499.8561396 MHz
DATA PROCESSING
Resol. enhancement -0.0 Hz
FT size 65536
Total time 0 min, 51 sec
7.274
7.166
6.885
6.883
4.266
4.148
4.130
4.124
4.101
4.085
4.083
3.520
3.505
3.408
3.233
3.227
3.222
3.202
3.188
3.180
3.178
3.152
3.029
3.025
2.01
3.08
3.38
2.07
2.57
ppm

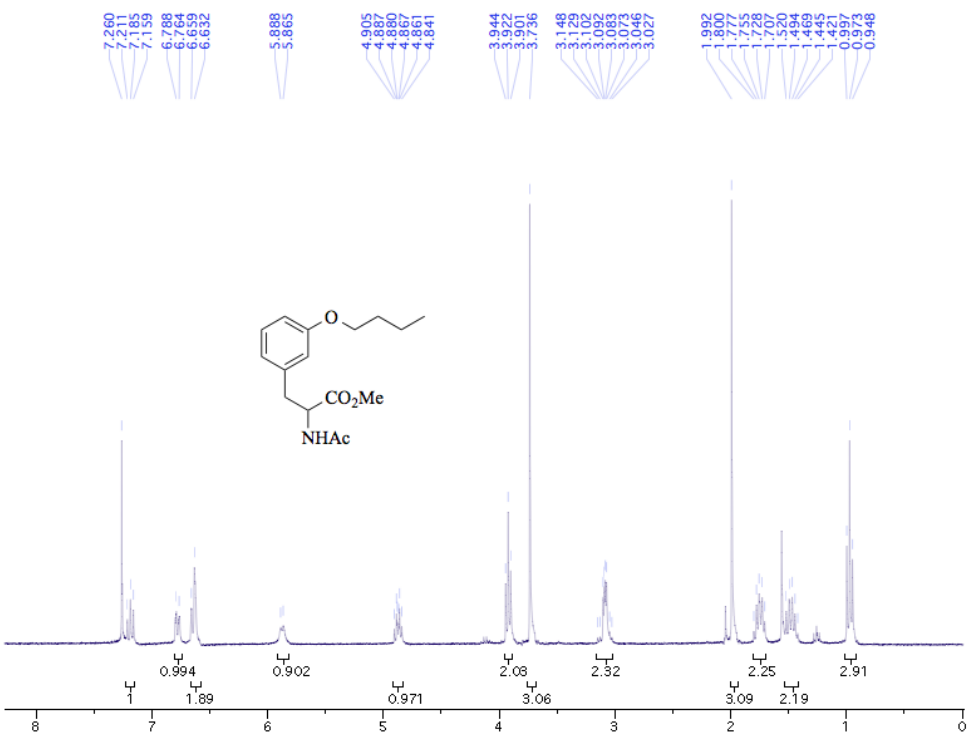

Std Carbon experiment
File: xp
Pulse Sequence: s2pul
Solvent: cdcl3
Ambient temperature
Operator: xfang
Mercury-300BB "mercury300"
Relax. delay 1.800 sec
Pulse 45.0 degrees
Acq. time 1.300 sec
Width 18083.2 Hz
512 repetitions
OBSERVE C13, 75.4138098 MHz
DECOUPLE H1, 299.9165608 MHz
Power 35 dB
Power 39 dB
continuously on
WALTZ-16 modulated
DATA PROCESSING
Line broadening 0.5 Hz
FT size 65536
Total time 28 min, 18 sec
ppm
180.287
158.113
128.488
114.789
85.028
67.007
59.813
49.876
48.420
48.184
47.856
47.571
47.285
47.000
46.722

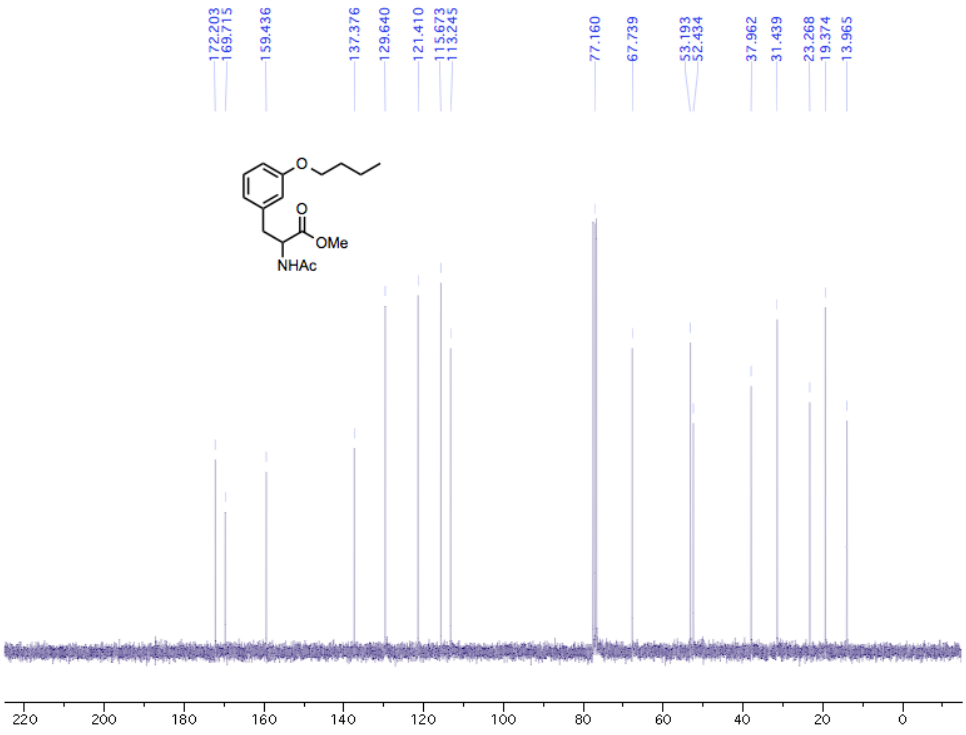

7.260
7.195
7.169
7.143
6.774
6.769
6.766
6.750
6.747
6.742
6.738
6.627
6.627
6.618
6.051
6.026
4.886
4.866
4.859
4.847
4.840
4.821
4.012
3.989
3.966
3.942
3.713
3.125
3.106
3.079
3.061
3.042
3.015
2.996
1.968
1.406
1.383
1.359
O
CO2Me
NHAc
0.961
0.889
2.3
2.58
3.39
1
2.03
1.05
3.33
3.32
ppm 10
9
8
7
6
5
4
3
2
1
0
-1

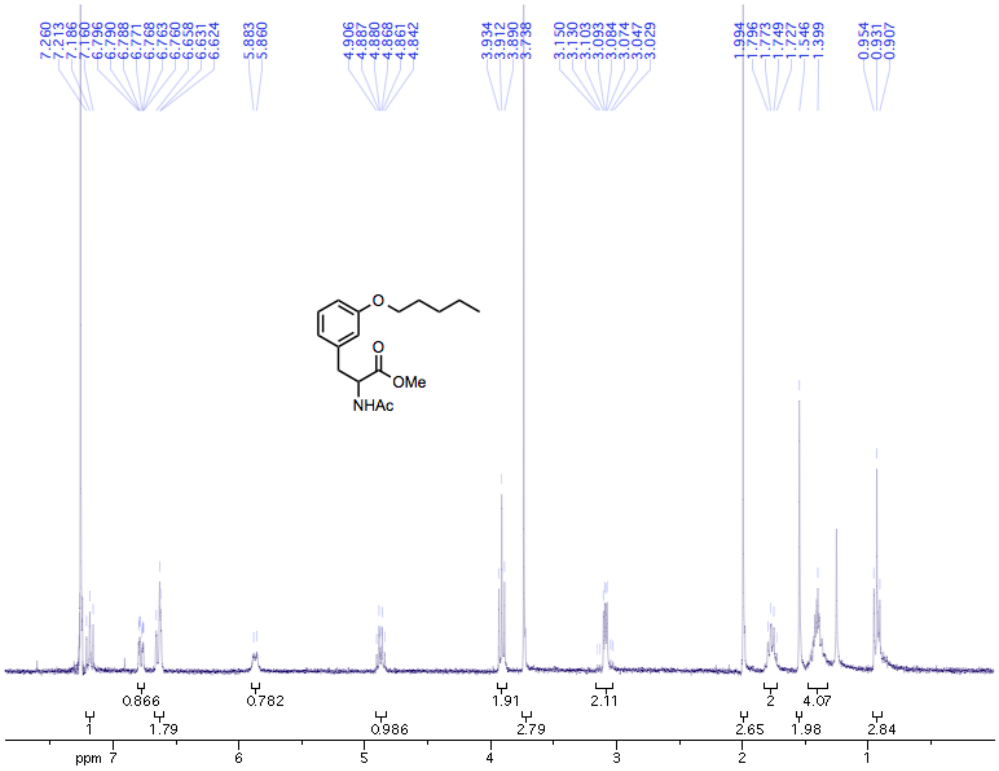

172.202
169.708
159.215
137.381
129.690
121.506
115.624
113.217
77.160
63.451
53.153
52.490
37.964
23.323
14.972
O
CO2Me
NHAc
80
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0

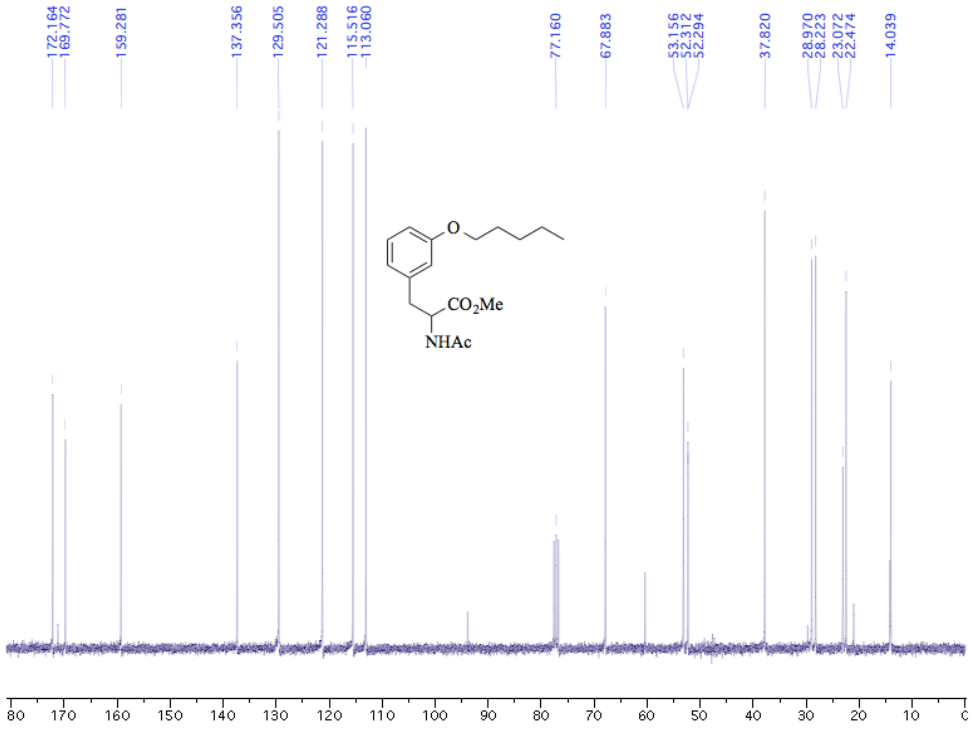

7.260
7.215
7.188
7.162
6.800
6.792
6.775
6.764
6.661
6.636
5.890
5.864
4.907
4.888
4.881
4.869
4.862
4.843
3.904
3.882
3.860
3.738
3.151
3.132
3.104
3.094
3.085
3.075
3.048
3.029
1.994
1.855
1.830
1.807
1.783
1.760
1.736
1.053
1.029
1.004
O
O
OMe
NHAc
1.01
0.87
2.06
2.25
2.28
1
1.97
1.02
2.95
2.98
3.11
9
8
7
6
5
4
3
2
1
0
-1
-2

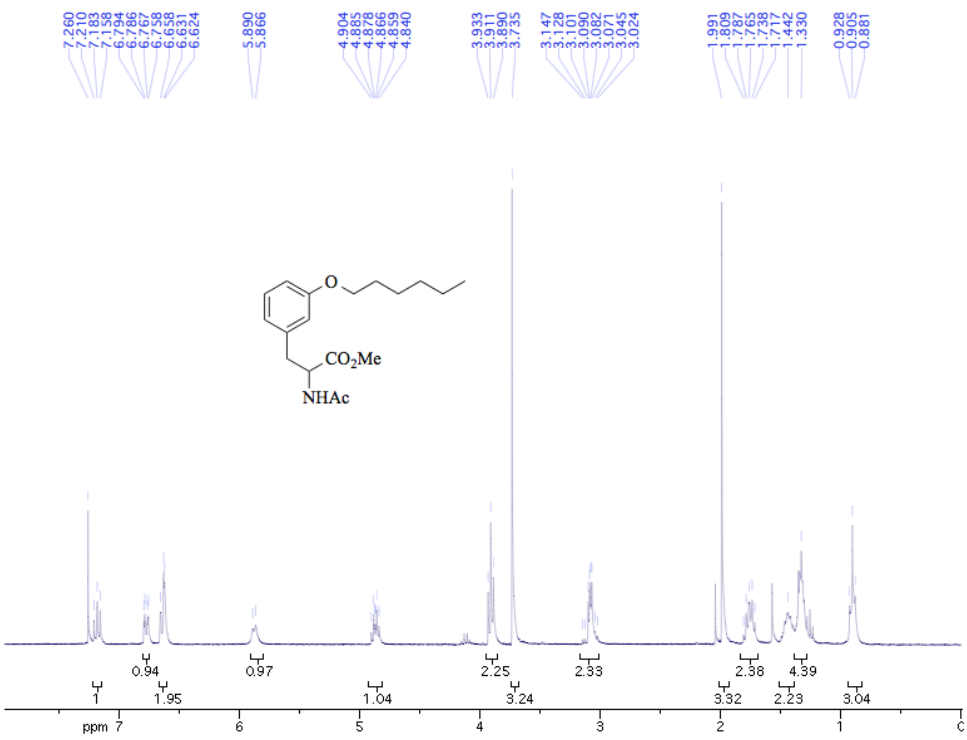

172.192
169.720
159.374
137.353
129.636
121.403
115.633
113.211
77.160
69.504
53.169
52.442
37.929
23.269
22.678
10.642
O
CO2Me
NHAc
80
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10

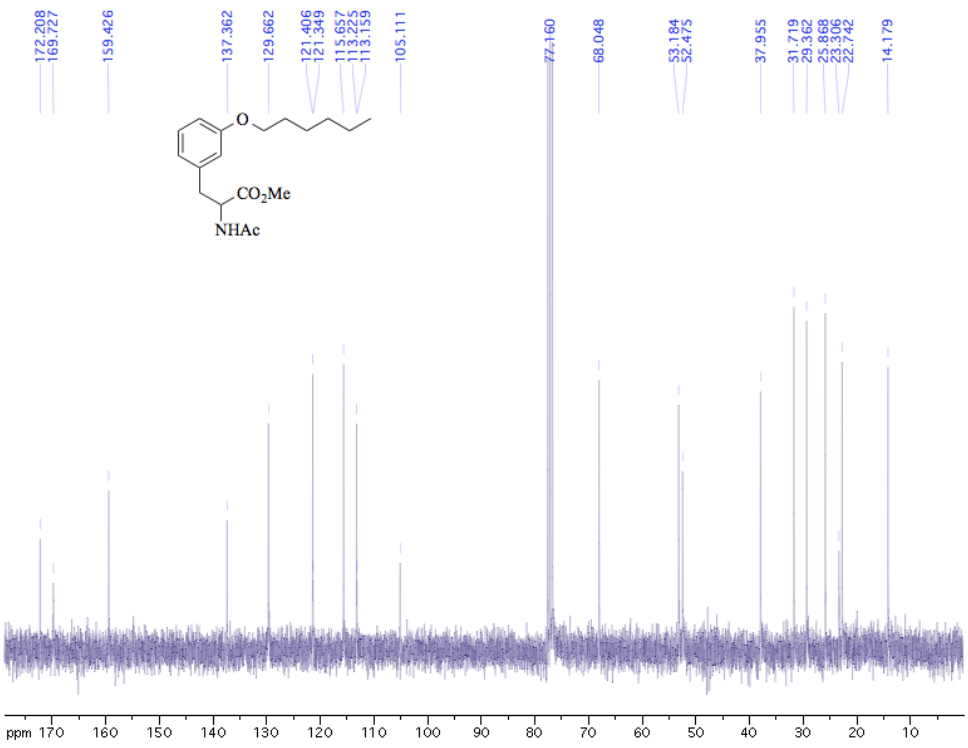

7.260
7.211
7.185
7.159
6.788
6.764
6.659
6.632
5.888
5.865
4.905
4.887
4.880
4.867
4.861
4.841
3.944
3.922
3.901
3.736
3.148
3.129
3.102
3.092
3.083
3.073
3.046
3.027
1.992
1.800
1.777
1.755
1.728
1.707
1.520
1.494
1.469
1.445
1.421
0.997
0.973
0.948
O
CO2Me
NHAc
0.994
0.902
2.08
2.32
2.25
2.91
1
1.89
0.971
3.06
3.09
2.19
8
7
6
5
4
3
2
1
0

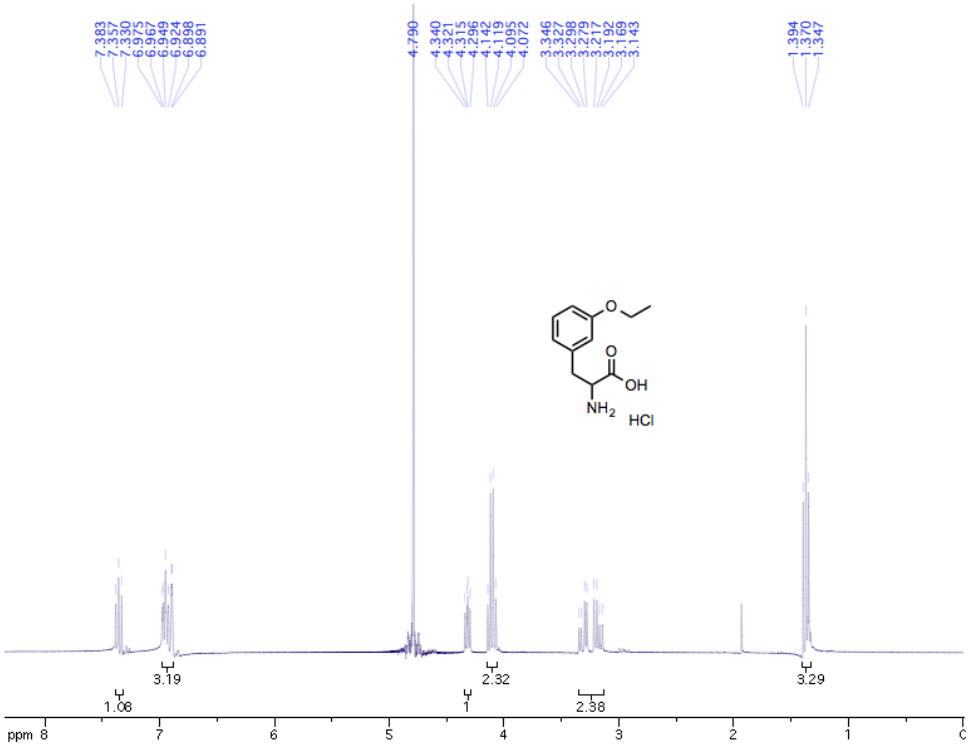

172.203
169.715
159.436
137.376
129.640
121.410
115.673
113.245
77.160
67.739
53.193
52.434
37.962
31.439
23.268
19.374
13.965
O
O
OMe
NHAc
220
200
180
160
140
120
100
80
60
40
20
0

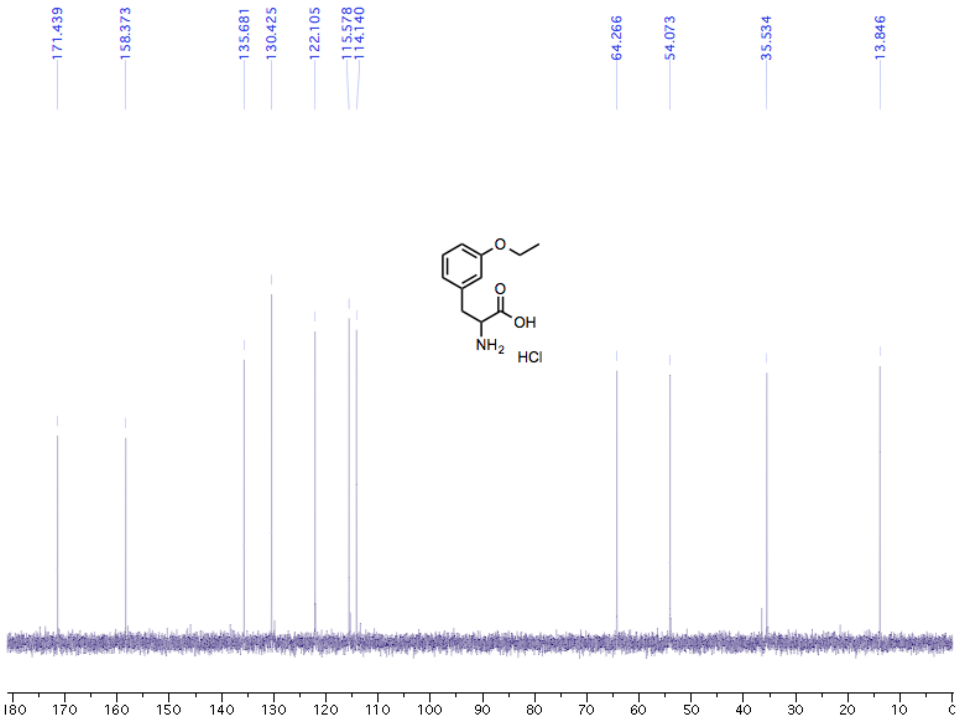

7.260
7.213
7.186
7.160
6.796
6.790
6.788
6.771
6.768
6.763
6.760
6.658
6.631
6.624
5.883
5.860
4.906
4.887
4.880
4.868
4.861
4.842
3.934
3.912
3.890
3.738
3.150
3.130
3.103
3.093
3.074
3.047
3.029
1.994
1.796
1.773
1.749
1.727
1.546
1.399
0.954
0.931
0.907
O
O
OMe
NHAc
0.866
0.782
1.91
2.11
2
4.07
1
1.79
0.986
2.79
2.65
1.98
2.84
ppm 7
6
5
4
3
2
1

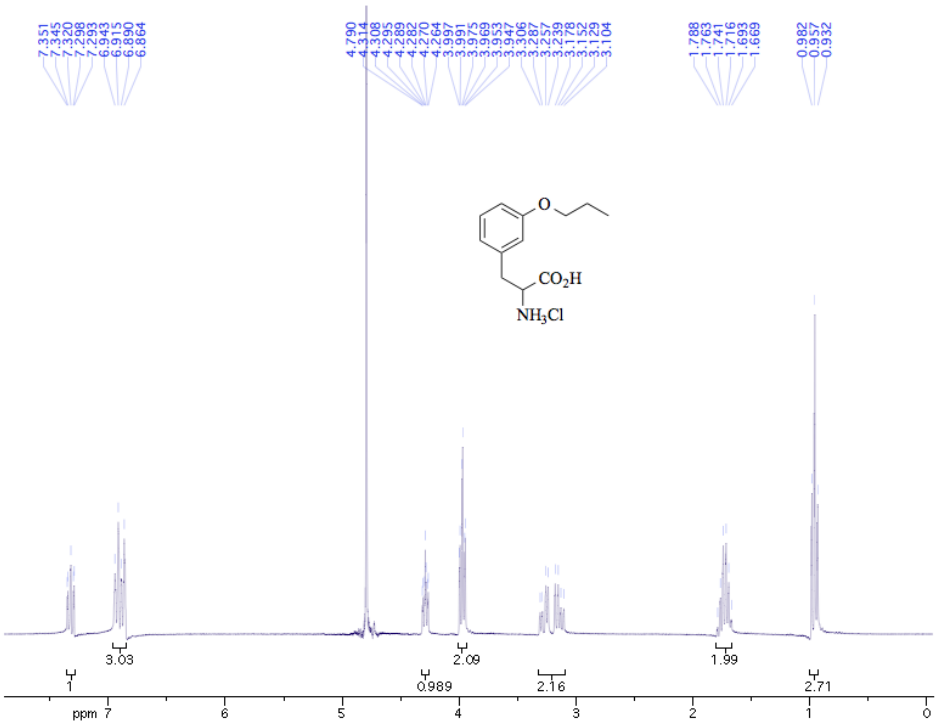

172.164
169.772
159.281
137.356
129.505
121.288
115.516
113.060
77.160
67.883
53.156
52.312
52.294
37.820
28.970
28.223
23.072
22.474
14.039
O
CO2Me
NHAc
80
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0

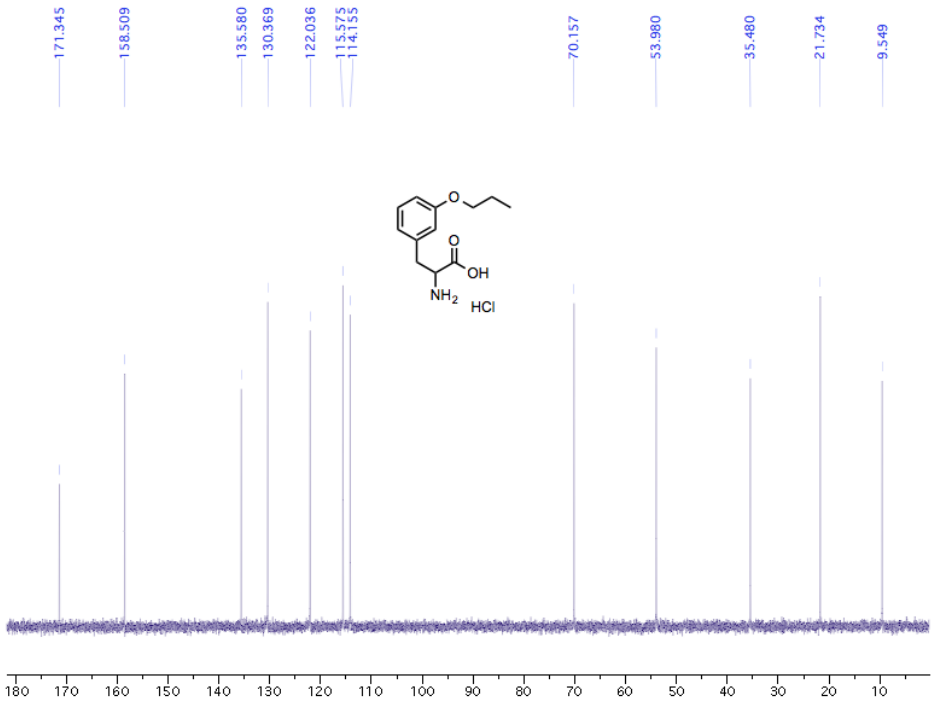

7.260
7.210
7.183
7.158
6.794
6.786
6.767
6.758
6.658
6.631
6.624
5.890
5.866
4.904
4.885
4.878
4.866
4.859
4.840
3.933
3.911
3.890
3.735
3.147
3.128
3.101
3.090
3.082
3.071
3.045
3.024
1.991
1.809
1.787
1.765
1.738
1.717
1.442
1.330
0.928
0.905
0.881
O
CO2Me
NHAc
0.94
0.97
2.25
2.33
2.38
4.39
1
1.95
1.04
3.24
3.32
2.23
3.04
ppm 7
6
5
4
3
2
1
0

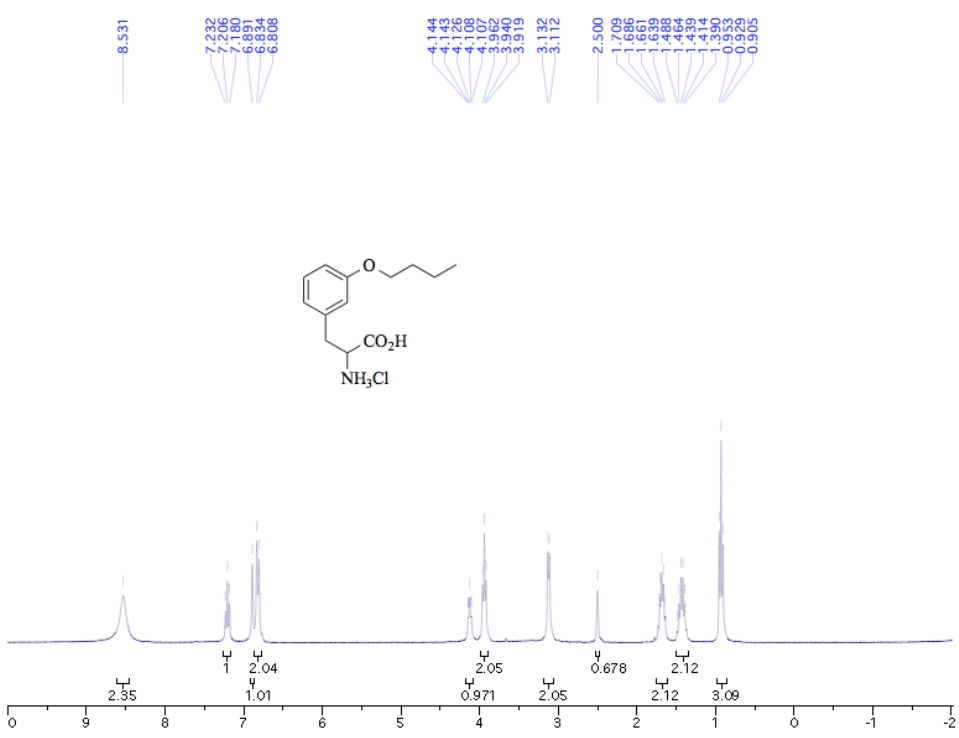

172.208
169.727
159.426
137.362
129.662
121.406
121.349
115.657
113.225
113.159
105.111
77.160
68.048
53.184
52.475
37.955
31.719
29.362
25.868
23.306
22.742
14.179
O
CO2Me
NHAc
ppm 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10

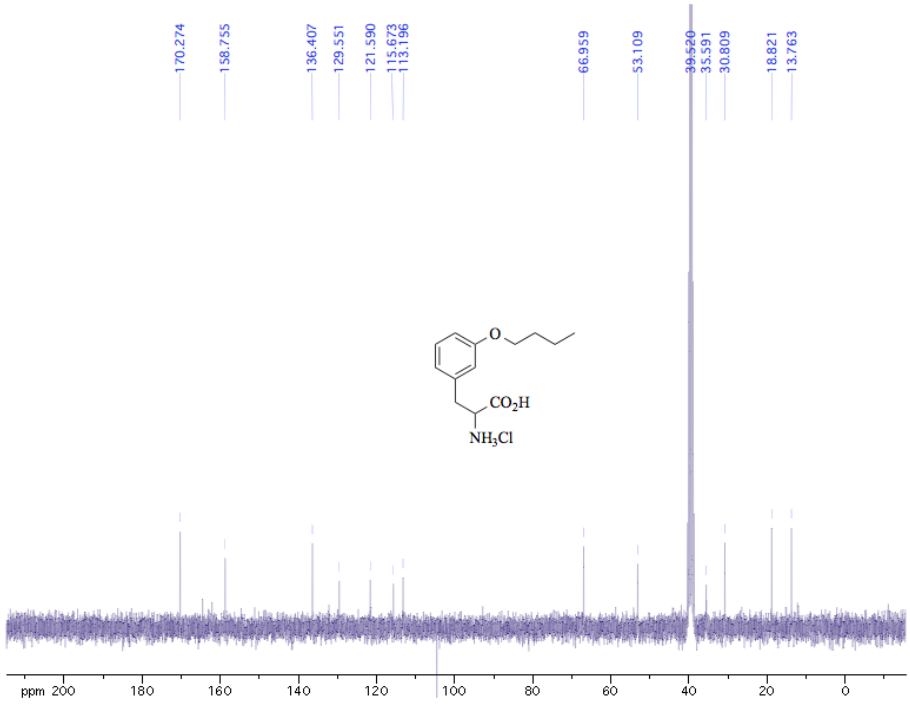

7.383
7.357
7.330
6.975
6.967
6.949
6.924
6.898
6.891
4.790
4.340
4.321
4.315
4.296
4.142
4.119
4.095
4.072
3.346
3.327
3.298
3.279
3.217
3.192
3.169
3.143
1.394
1.370
1.347
O
O
OH
NH2
HCl
3.19
2.32
3.29
1.08
1
2.38
ppm 8
7
6
5
4
3
2
1
0

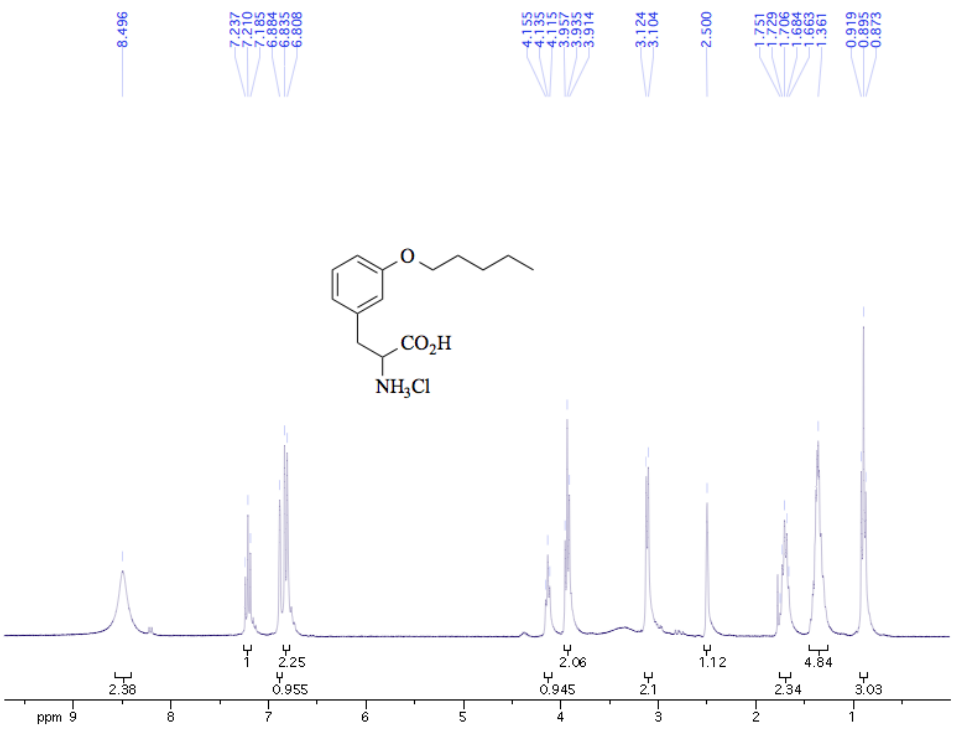

171.439
158.373
135.681
130.425
122.105
115.578
114.140
64.266
54.073
35.534
13.846
O
O
OH
NH2
HCl
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0

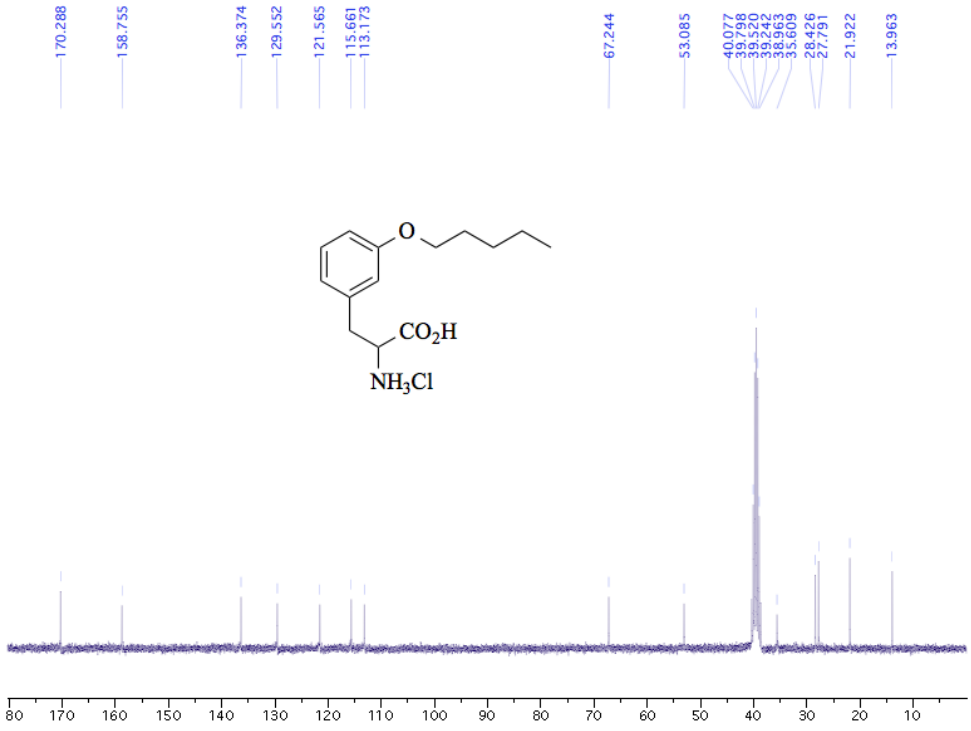

7.351
7.345
7.320
7.298
7.293
6.943
6.915
6.890
6.864
4.790
4.314
4.308
4.285
4.282
4.270
4.264
3.997
3.991
3.975
3.969
3.953
3.947
3.306
3.287
3.257
3.239
3.178
3.152
3.129
3.104
1.788
1.763
1.741
1.716
1.693
1.669
0.982
0.957
0.932
O
CO2H
NH3Cl
3.03
2.09
1.99
1
0.989
2.16
2.71
ppm 7
6
5
4
3
2
1
0

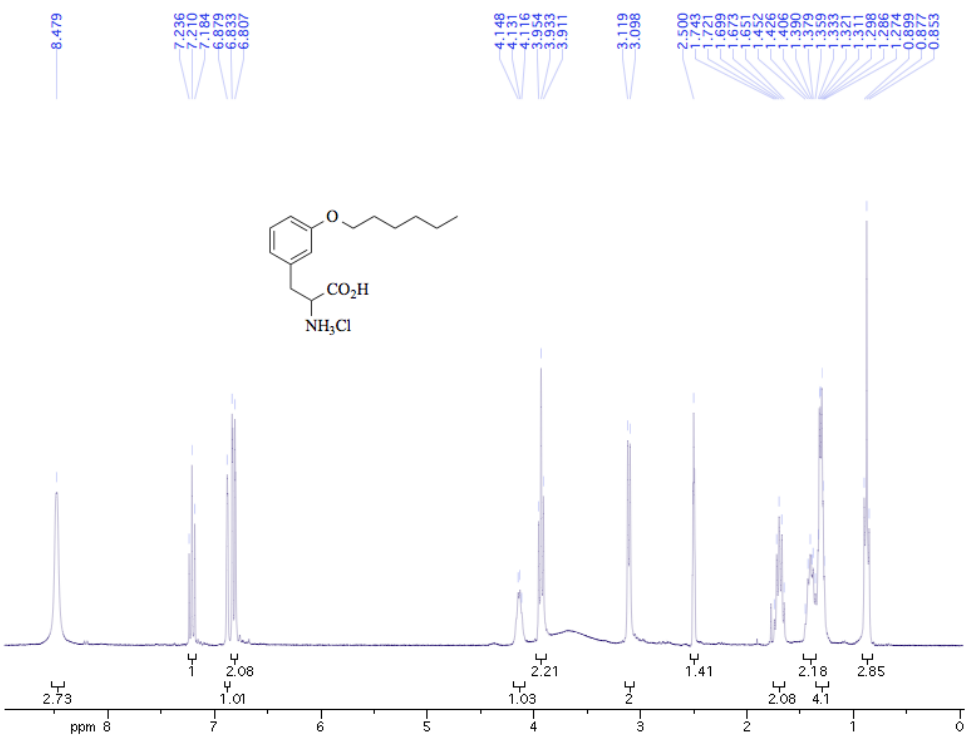

171.345
158.509
135.580
130.369
122.036
115.575
114.155
70.157
53.980
35.480
21.734
9.549
O
O
OH
NH2
HCl
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10

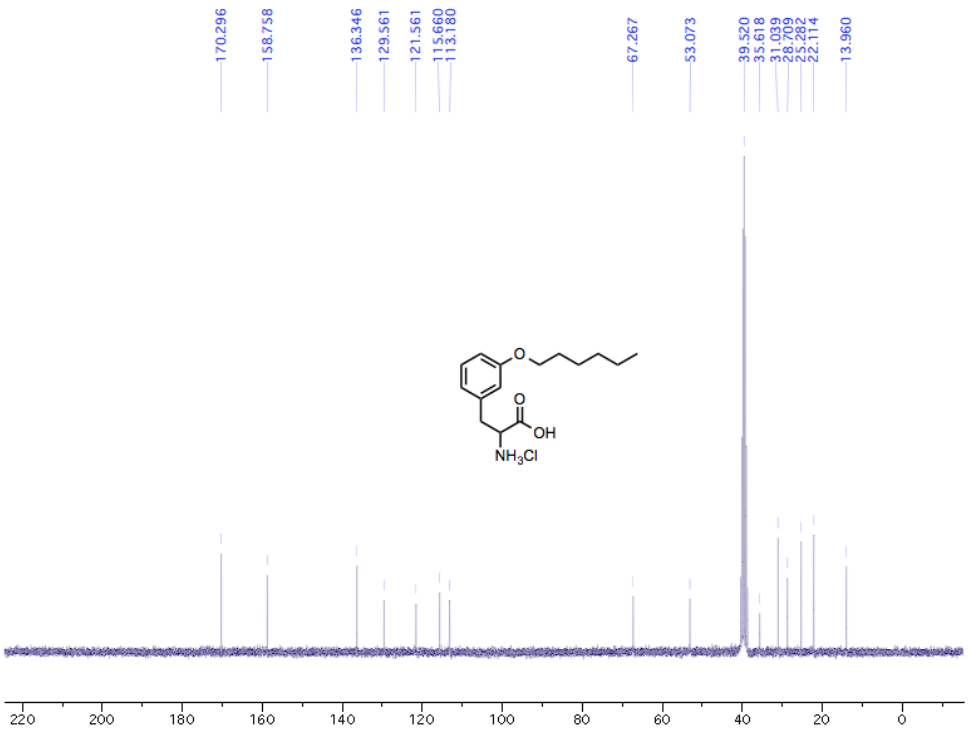

8.531
7.232
7.206
7.180
6.891
6.834
6.808
4.144
4.143
4.126
4.108
4.107
3.962
3.940
3.919
3.132
3.112
2.500
1.709
1.686
1.661
1.639
1.488
1.464
1.439
1.414
1.390
0.953
0.929
0.905
O
CO2H
NH3Cl
1
2.04
2.05
0.678
2.12
2.35
1.01
0.971
2.05
2.12
3.09
0
9
8
7
6
5
4
3
2
1
0
-1
-2

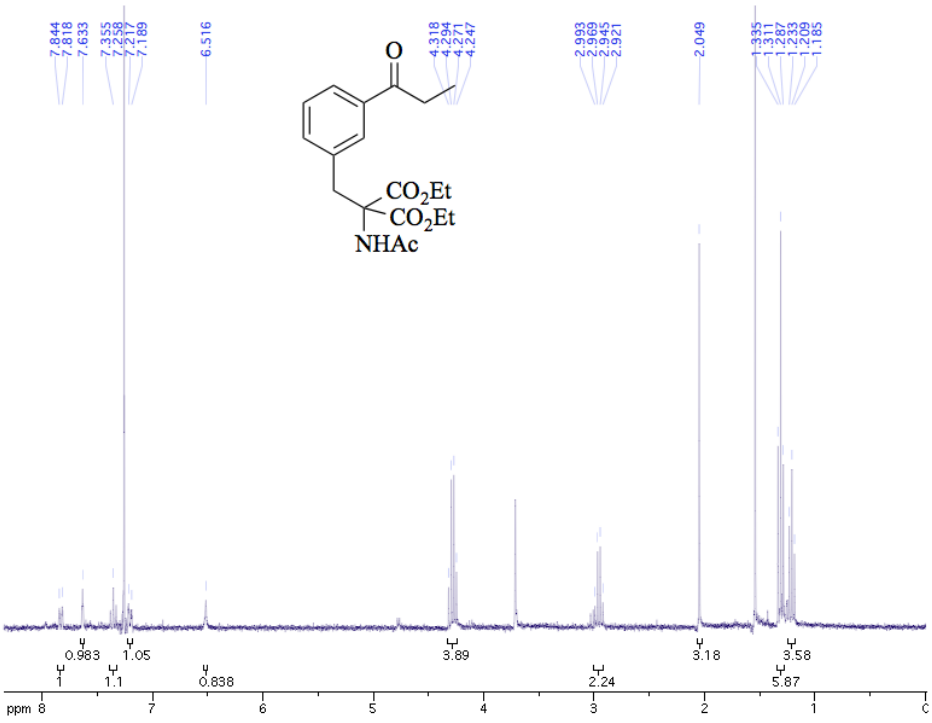

170.274
158.755
136.407
129.551
121.590
115.673
113.196
66.959
53.109
39.520
35.591
30.809
18.821
13.763
O
CO2H
NH3Cl
ppm 200
180
160
140
120
100
80
60
40
20
0

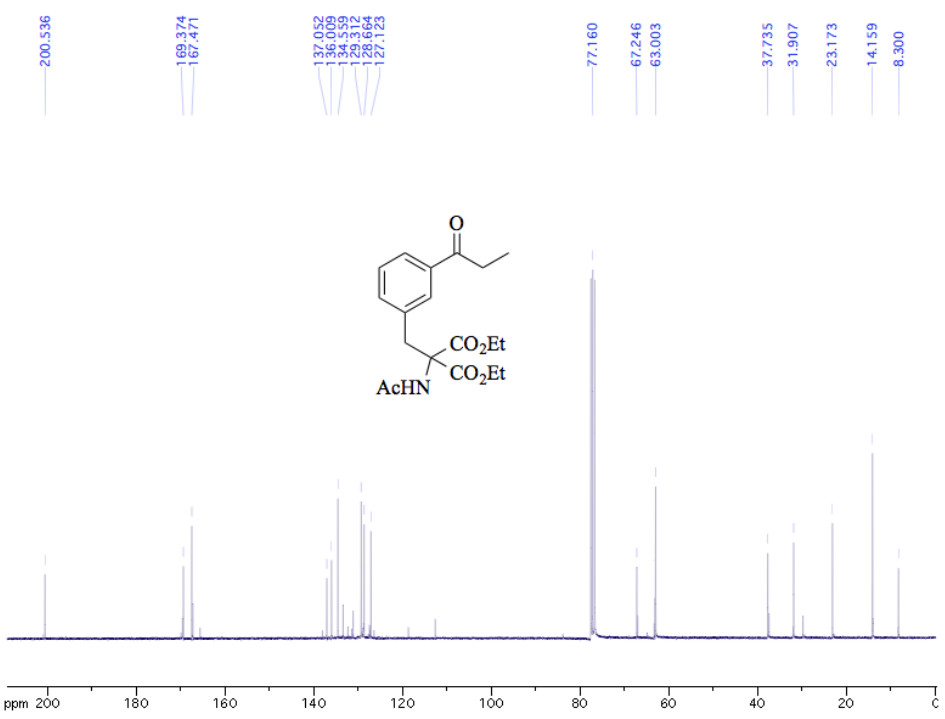

8.496
7.237
7.210
7.185
6.884
6.835
6.808
4.155
4.135
4.115
3.957
3.935
3.914
3.124
3.104
2.500
1.751
1.729
1.706
1.684
1.663
1.561
0.919
0.895
0.873
O
CO2H
NH3Cl
1
2.25
2.06
1.12
4.84
2.38
0.955
0.945
2.1
2.34
3.03
ppm 9
8
7
6
5
4
3
2
1

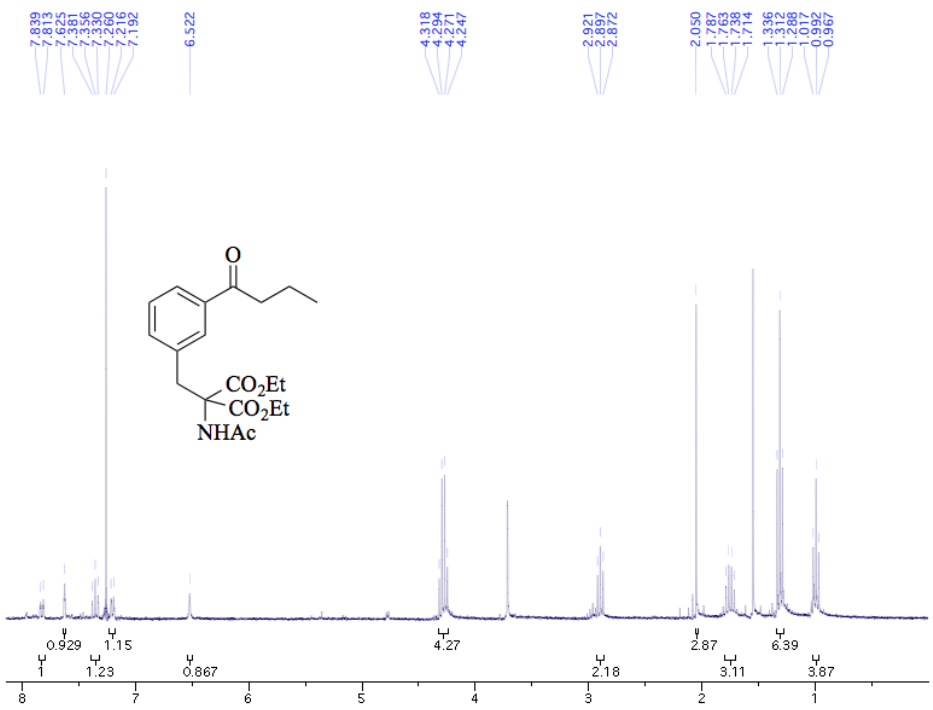

170.288
158.755
136.374
129.552
121.565
115.661
113.173
67.244
53.085
40.077
39.758
39.440
39.242
38.863
35.609
28.426
27.791
21.922
13.963
O
CO2H
NH3Cl
80
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10

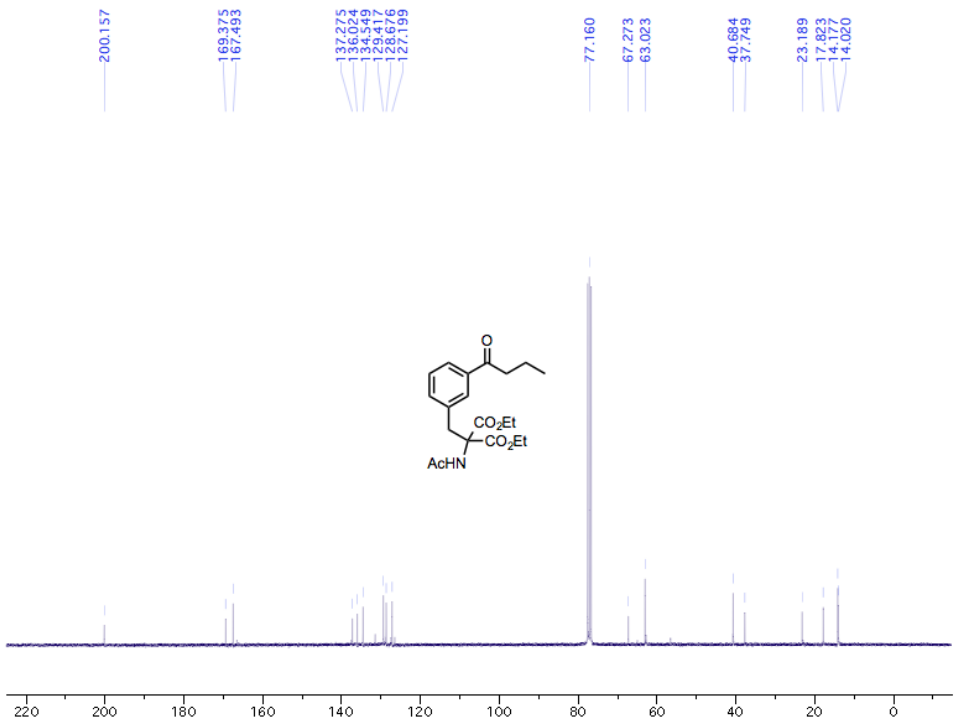

8.479
7.236
7.210
7.184
6.879
6.833
6.807
4.148
4.131
4.116
3.954
3.933
3.911
3.119
3.098
2.500
2.343
1.721
1.699
1.673
1.651
1.452
1.426
1.406
1.390
1.379
1.359
1.333
1.321
1.311
1.298
1.286
1.274
0.899
0.877
0.853
O
CO2H
NH3Cl
1
2.08
2.21
1.41
2.18
2.85
2.73
1.01
1.03
2
2.08
4.1
ppm 8
7
6
5
4
3
2
1
0

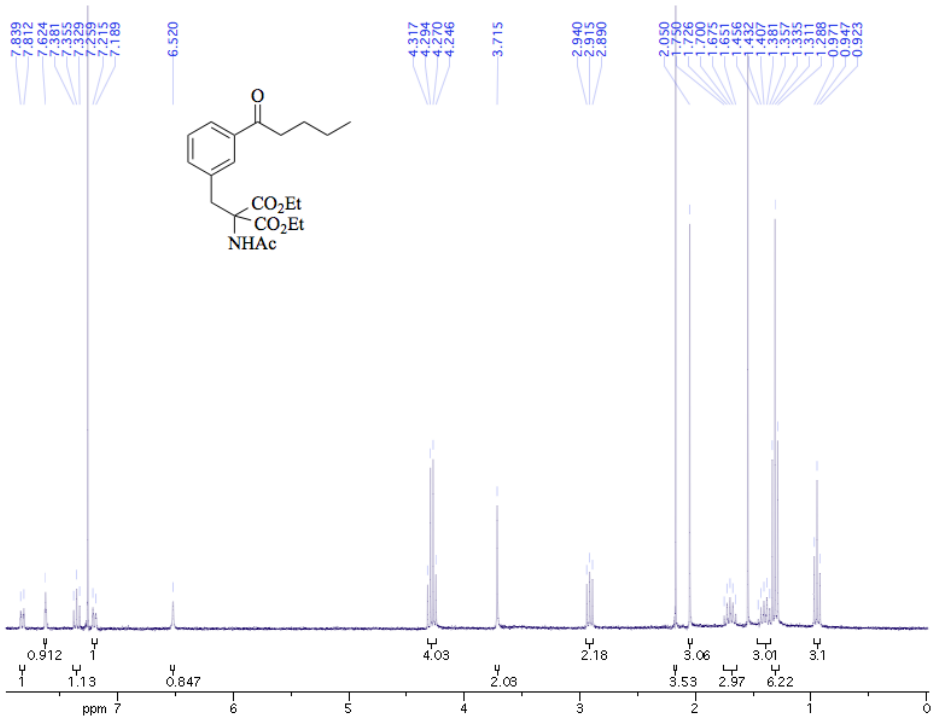

170.296
158.758
136.346
129.561
121.561
115.660
113.180
67.267
53.073
39.520
35.618
31.039
28.709
25.282
22.114
13.960
O
O
OH
NH3Cl
220
200
180
160
140
120
100
80
60
40
20
0

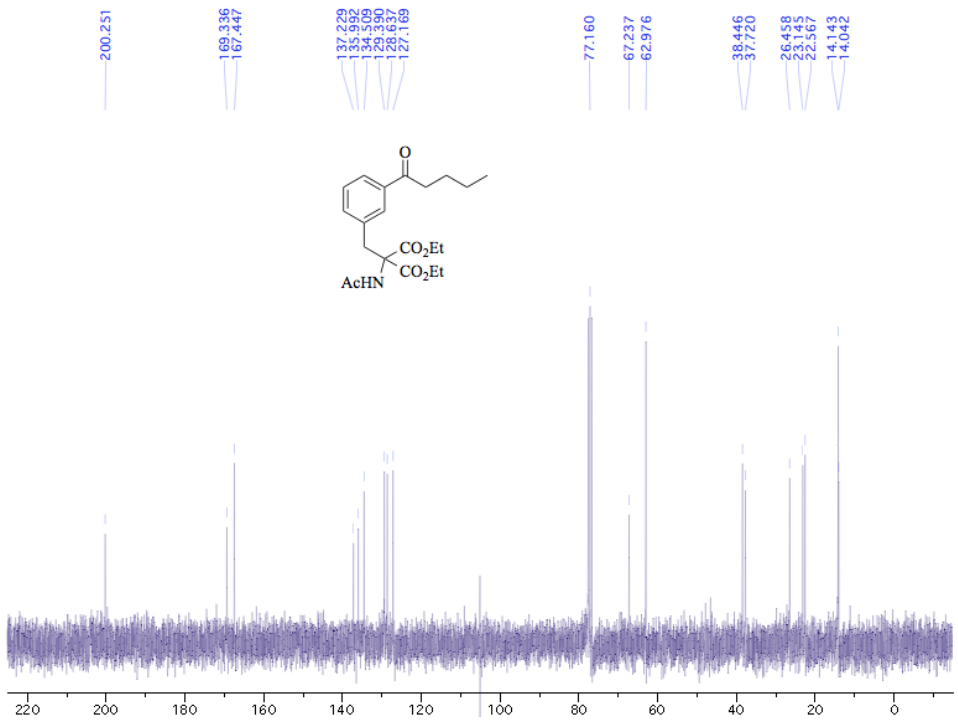

7.844
7.818
7.633
7.355
7.258
7.217
7.189
6.516
4.318
4.294
4.271
4.247
2.993
2.969
2.945
2.921
2.049
1.335
1.311
1.287
1.263
1.209
1.185
O
CO2Et
CO2Et
NHAc
0.983
1.05
3.89
3.18
3.58
1
1.1
0.838
2.24
5.87
ppm 8
7
6
5
4
3
2
1
0

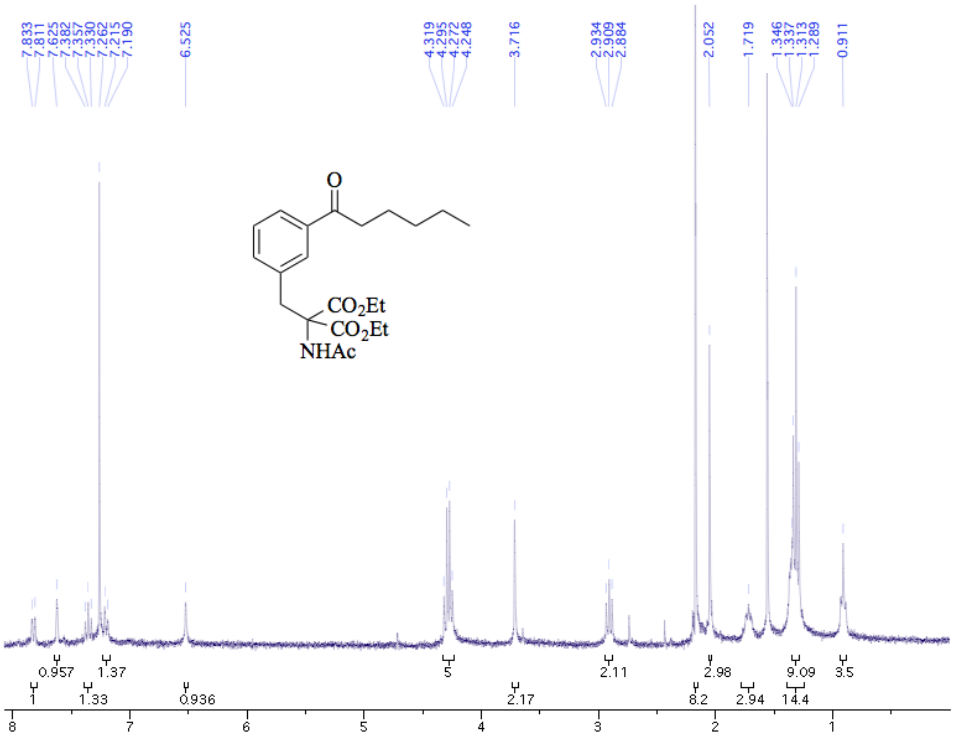

200.536
169.374
167.471
137.052
136.009
134.559
129.312
128.664
127.123
77.160
67.246
63.003
37.735
31.907
23.173
14.159
8.300
O
CO2Et
CO2Et
AcHN
ppm 200
180
160
140
120
100
80
60
40
20
0

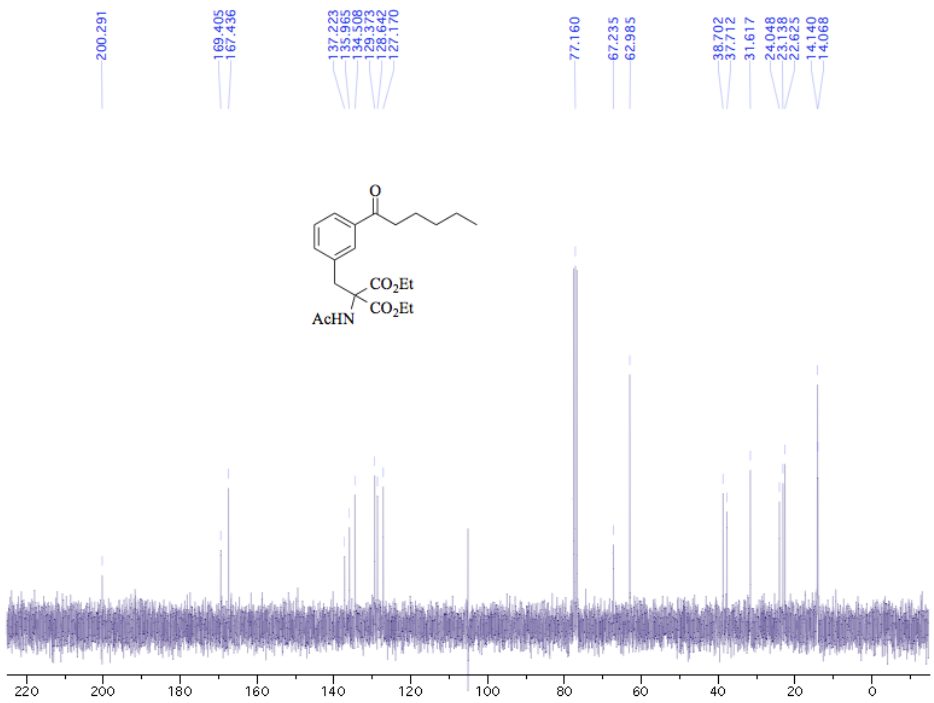

7.839
7.813
7.625
7.381
7.356
7.330
7.260
7.216
7.192
6.522
4.318
4.294
4.271
4.247
2.921
2.897
2.872
2.050
1.787
1.763
1.738
1.714
1.336
1.312
1.288
1.017
0.992
0.967
O
CO2Et
CO2Et
NHAc
0.929
1.15
4.27
2.87
6.39
1
1.23
0.867
2.18
3.11
3.87
8
7
6
5
4
3
2
1

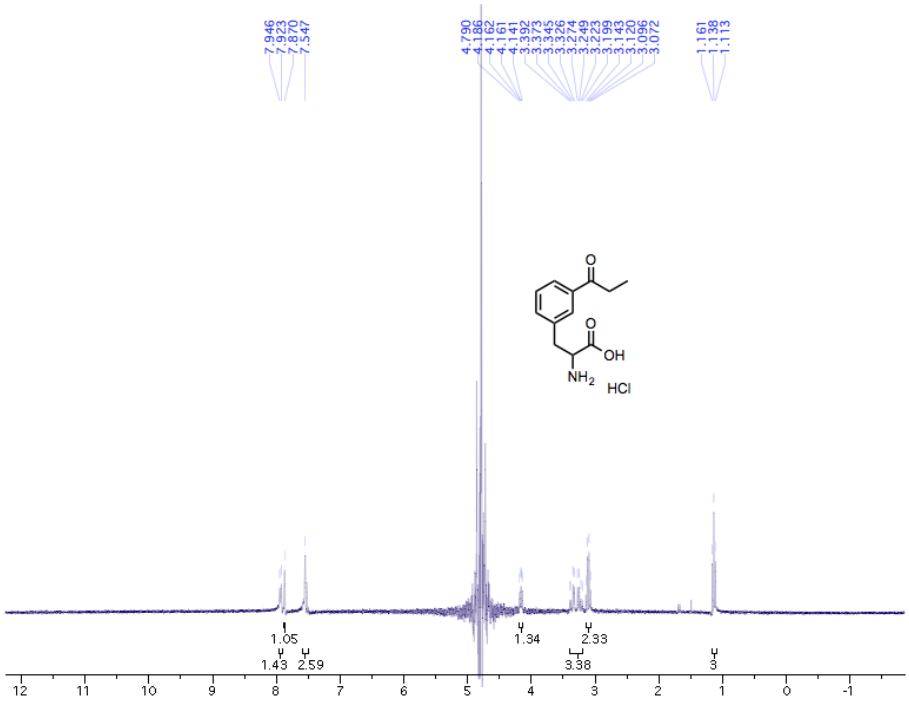

200.157
169.375
167.493
137.275
136.024
134.549
129.417
128.676
127.199
77.160
67.273
63.023
40.684
37.749
23.189
17.823
14.177
14.020
O
CO2Et
CO2Et
AcHN
220
200
180
160
140
120
100
80
60
40
20
0

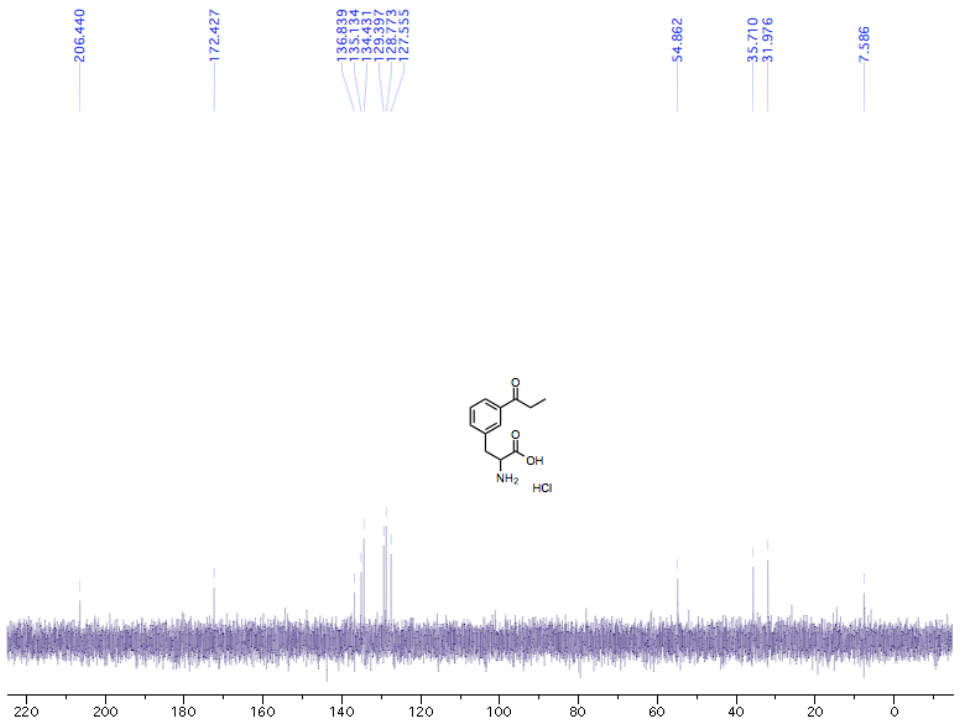

7.839
7.812
7.624
7.381
7.355
7.329
7.259
7.215
7.189
6.520
4.317
4.294
4.270
4.246
3.715
2.940
2.915
2.890
2.050
1.750
1.726
1.700
1.675
1.571
1.456
1.432
1.407
1.381
1.357
1.335
1.311
1.288
0.971
0.947
0.923
O
CO2Et
CO2Et
NHAc
0.912
1
4.03
2.18
3.06
3.01
3.1
1
1.13
0.847
2.08
3.53
2.97
6.22
ppm 7
6
5
4
3
2
1
0

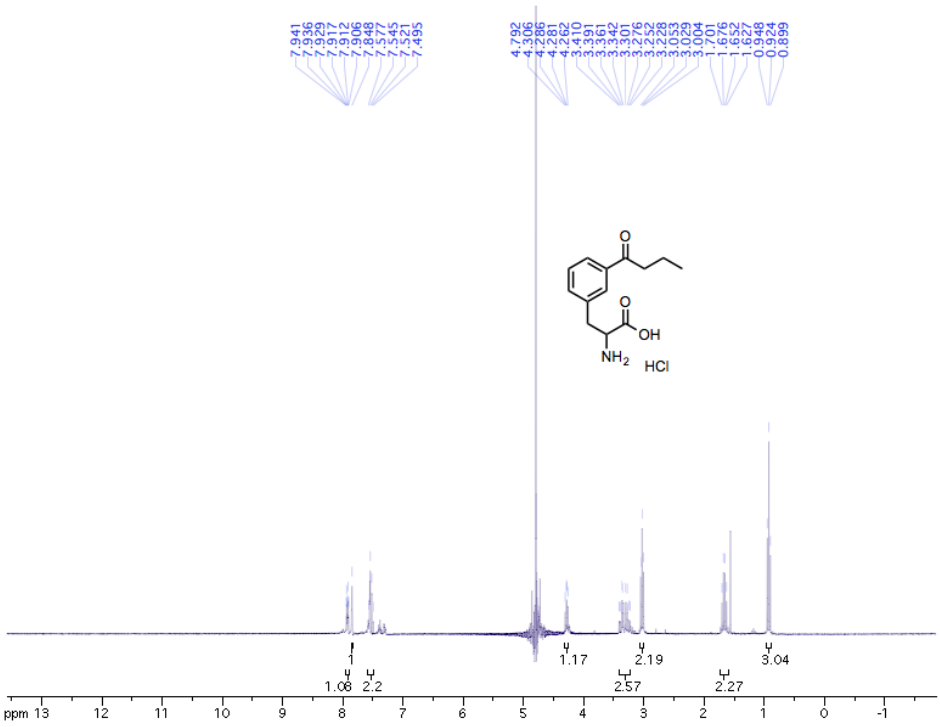

200.251
169.336
167.447
137.229
135.992
134.509
129.390
128.637
127.169
77.160
67.237
62.976
38.446
37.720
26.458
23.145
22.567
14.143
14.042
O
CO2Et
CO2Et
AcHN
220
200
180
160
140
120
100
80
60
40
20
0

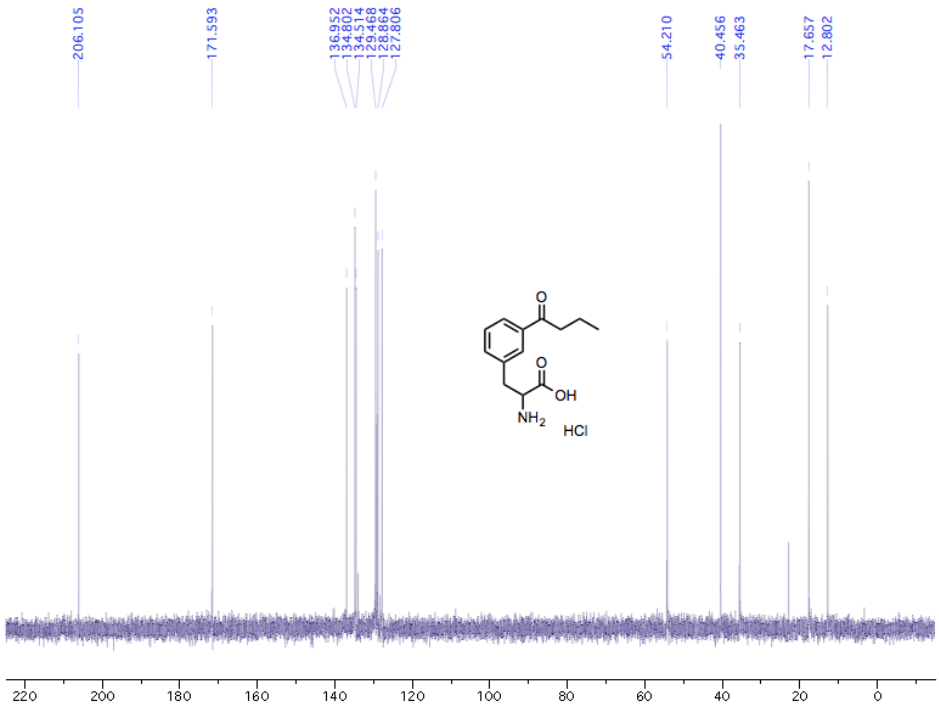

7.833
7.811
7.395
7.382
7.357
7.330
7.262
7.215
7.190
6.525
4.319
4.295
4.272
4.248
3.716
2.934
2.909
2.884
2.052
1.719
1.346
1.337
1.313
1.289
0.911
O
CO2Et
CO2Et
NHAc
0.957
1.37
5
2.11
2.98
9.09
3.5
1
1.33
0.936
2.17
8.2
2.94
14.4
8
7
6
5
4
3
2
1

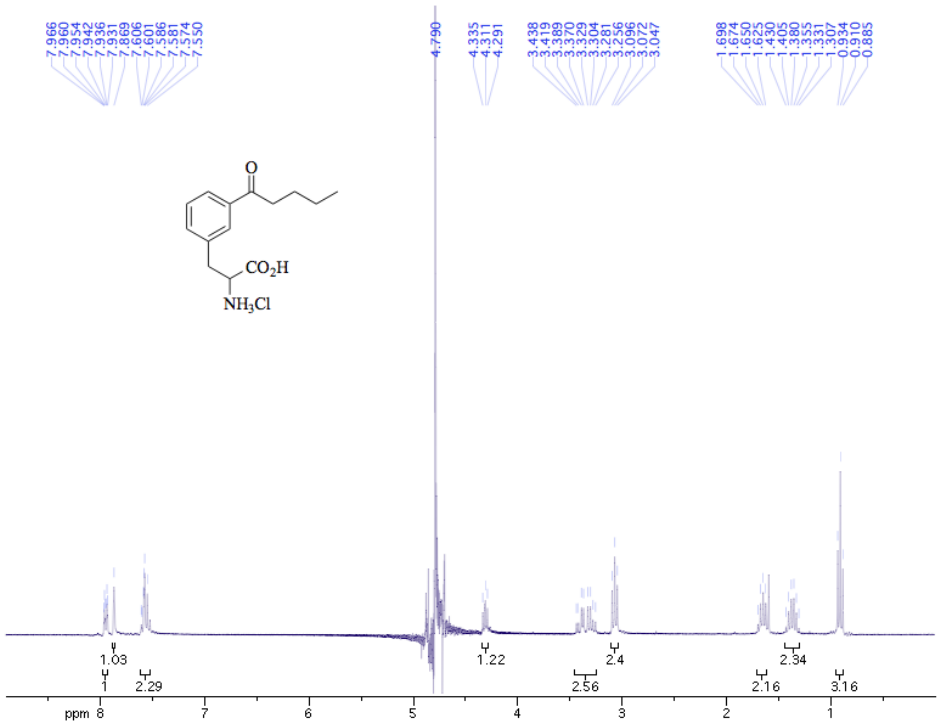

200.291
169.405
167.436
137.223
135.965
134.508
129.373
128.642
127.170
77.160
67.235
62.985
38.702
37.712
31.617
24.048
23.138
22.425
14.140
14.068
O
CO2Et
CO2Et
AcHN
220
200
180
160
140
120
100
80
60
40
20
0

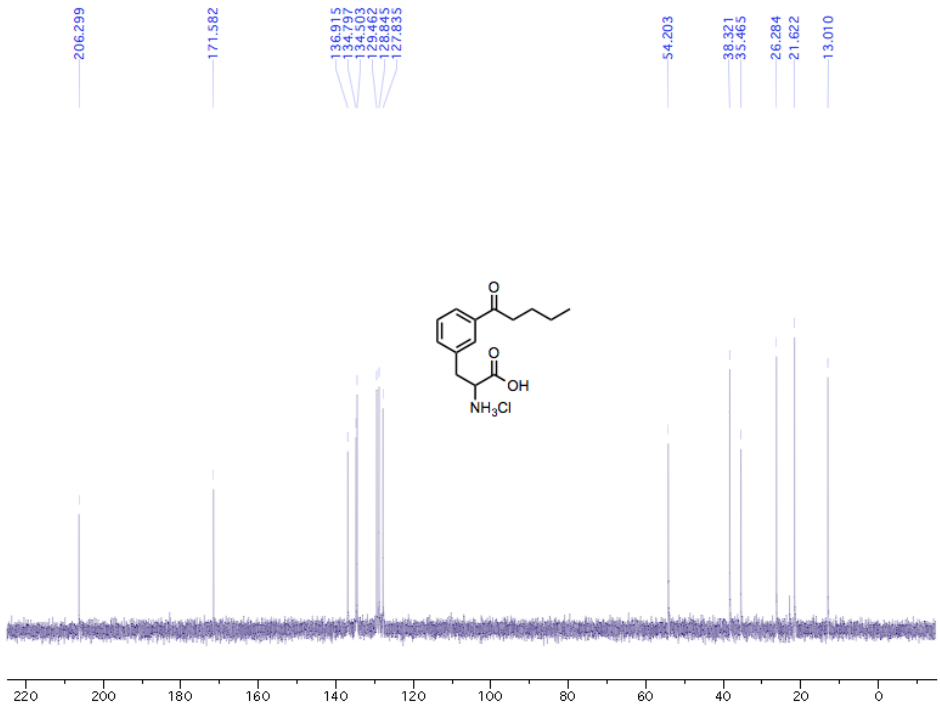

7.946
7.923
7.870
7.347
4.790
4.186
4.162
4.161
4.141
3.392
3.373
3.345
3.326
3.274
3.249
3.223
3.199
3.143
3.120
3.096
3.072
1.161
1.138
1.113
O
O
OH
NH2
HCl
1.05
1.43
2.59
1.34
2.33
3.38
3
12
11
10
9
8
7
6
5
4
3
2
1
0
-1

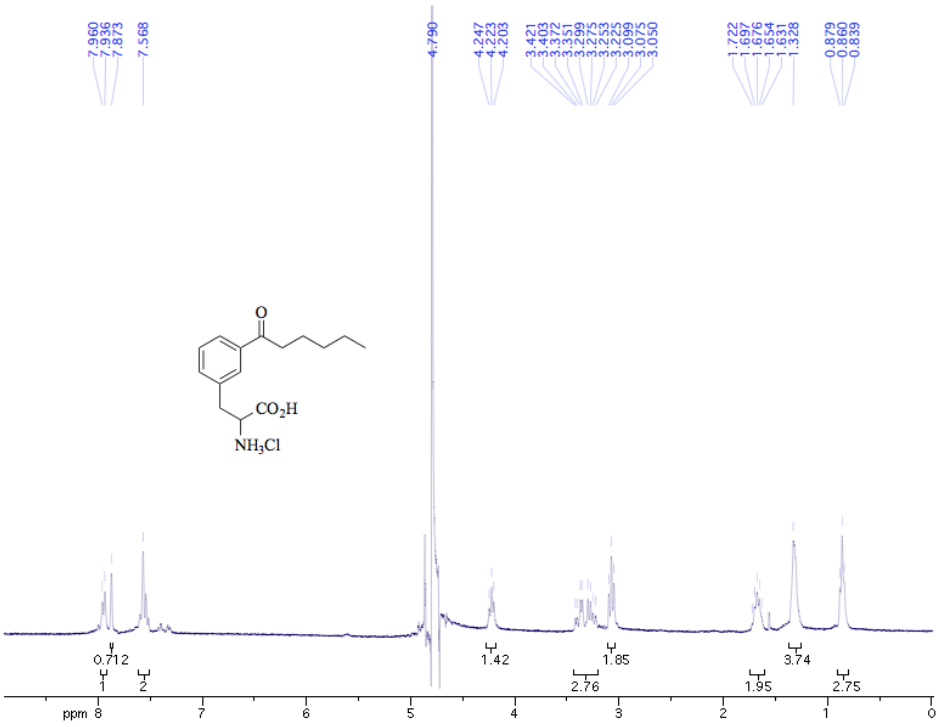

206.440
172.427
136.839
135.134
134.431
129.397
129.273
127.555
54.862
35.710
31.976
7.586
O
O
OH
NH2
HCl
220
200
180
160
140
120
100
80
60
40
20
0

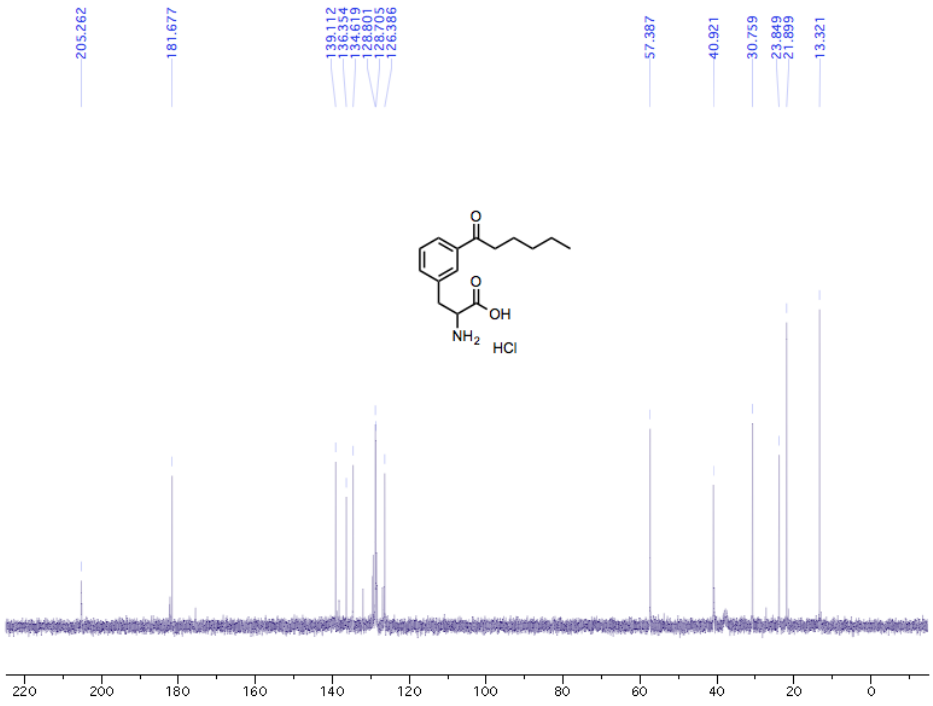

7.941
7.936
7.929
7.917
7.912
7.906
7.848
7.577
7.545
7.521
7.495
4.792
4.306
4.286
4.281
4.262
3.410
3.391
3.361
3.342
3.301
3.276
3.252
3.228
3.053
3.029
3.004
1.701
1.676
1.652
1.627
0.948
0.924
0.899
O
O
OH
NH2
HCl
1
1.08
2.2
1.17
2.19
2.57
3.04
2.27
ppm 13
12
11
10
9
8
7
6
5
4
3
2
1
0
-1


206.105
171.593
136.952
134.902
134.514
129.468
128.864
127.806
54.210
40.456
35.463
17.657
12.802
O
O
OH
NH2
HCl
220
200
180
160
140
120
100
80
60
40
20
0


7.966
7.960
7.954
7.942
7.936
7.931
7.669
7.666
7.601
7.586
7.581
7.574
7.550
4.790
4.335
4.311
4.291
3.438
3.419
3.389
3.370
3.329
3.304
3.281
3.256
3.096
3.072
3.047
1.698
1.674
1.650
1.625
1.430
1.405
1.380
1.355
1.331
1.307
0.934
0.910
0.885
O
CO2H
NH3Cl
1.03
2.29
1.22
2.4
2.56
2.34
2.16
3.16
ppm 8
7
6
5
4
3
2
1

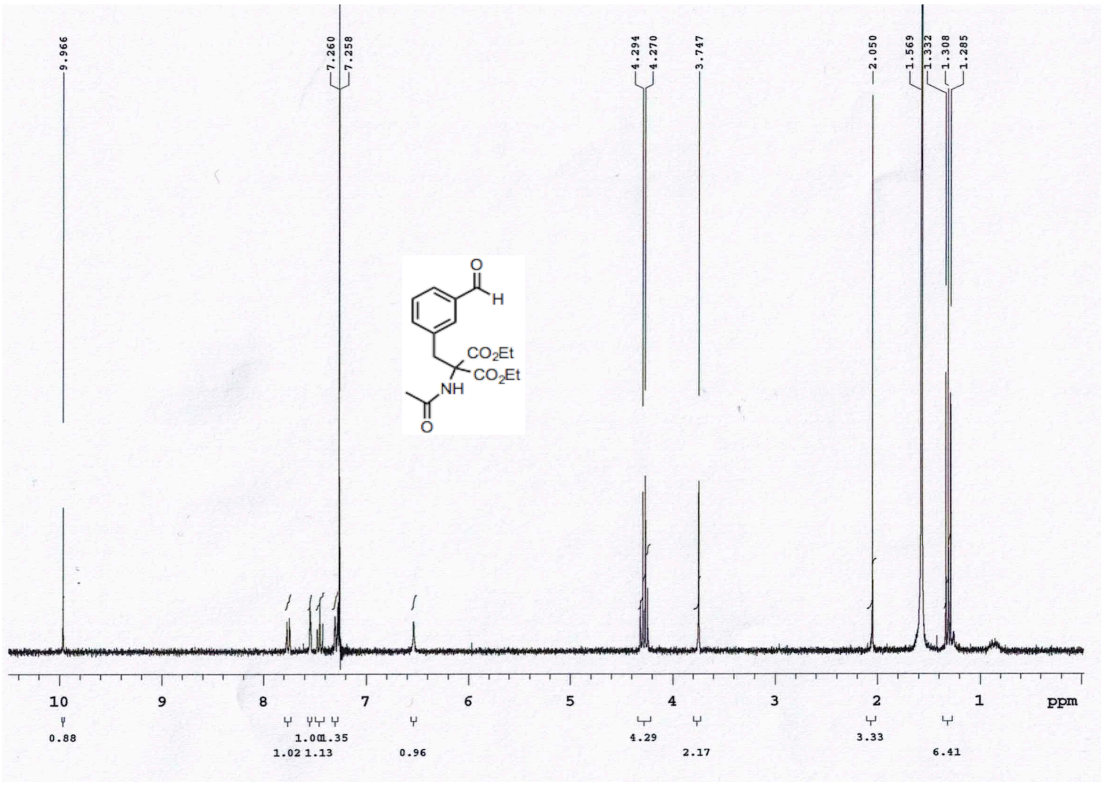

206.299
171.582
136.915
134.797
134.503
129.462
128.845
127.835
54.203
38.321
35.465
26.284
21.622
13.010
O
O
OH
NH3Cl
220
200
180
160
140
120
100
80
60
40
20
0

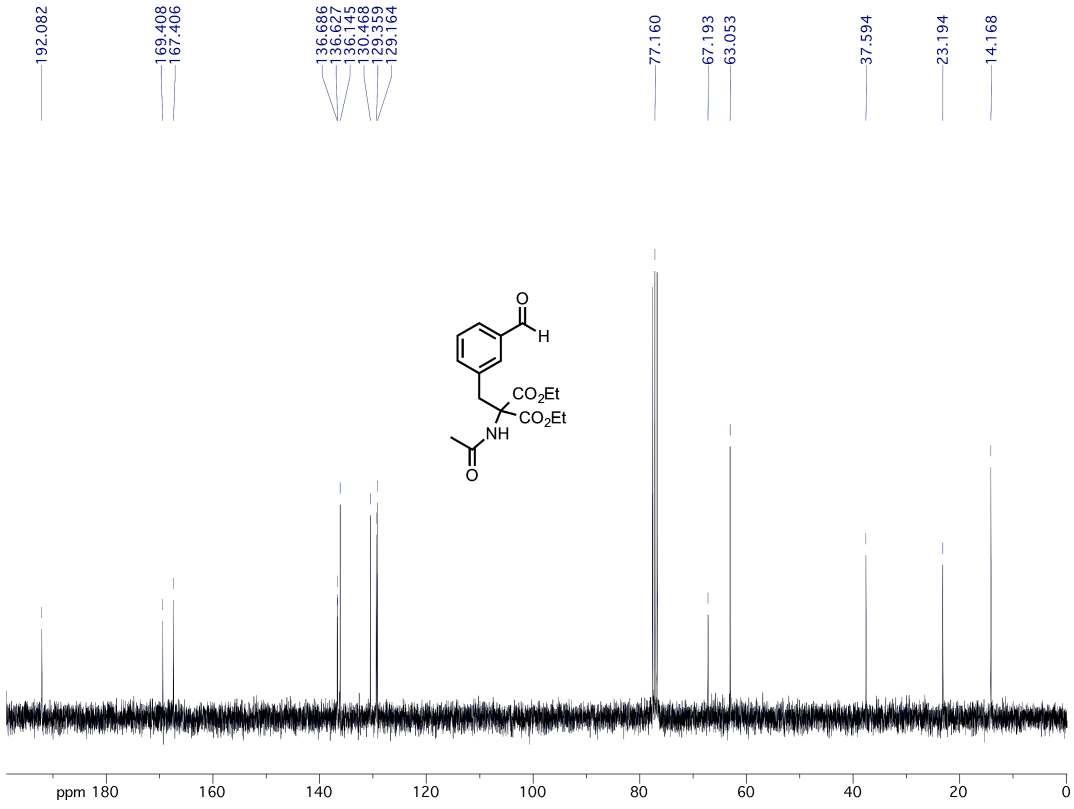

7.960
7.936
7.873
7.568
4.790
4.247
4.223
4.203
3.421
3.403
3.372
3.351
3.299
3.275
3.253
3.225
3.099
3.075
3.050
1.722
1.697
1.676
1.654
1.631
1.328
0.879
0.860
0.839
O
CO2H
NH3Cl
0.712
1
2
1.42
2.76
1.85
1.95
3.74
2.75
ppm 8
7
6
5
4
3
2
1
0

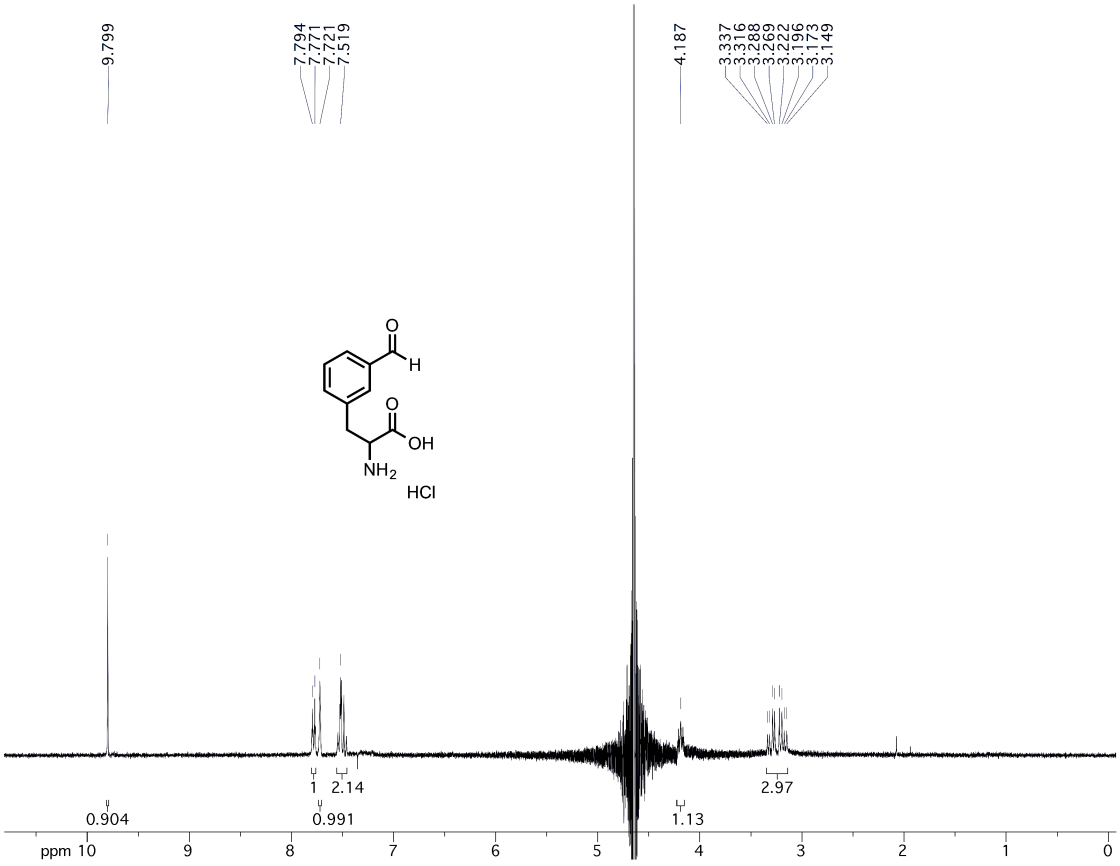

205.262
181.677
139.112
136.354
134.619
128.801
128.705
126.386
57.387
40.921
30.759
23.849
21.899
13.321
O
O
OH
NH2
HCl
220
200
180
160
140
120
100
80
60
40
20
0

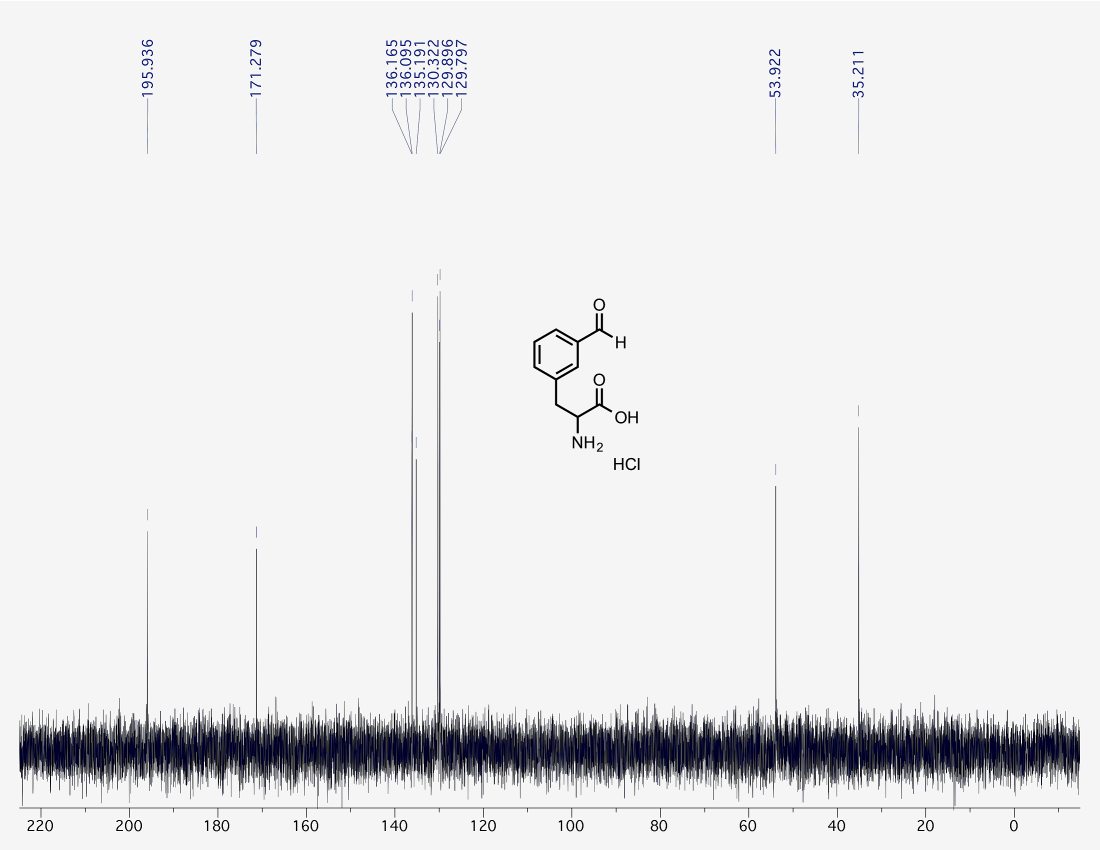
APPENDIX B

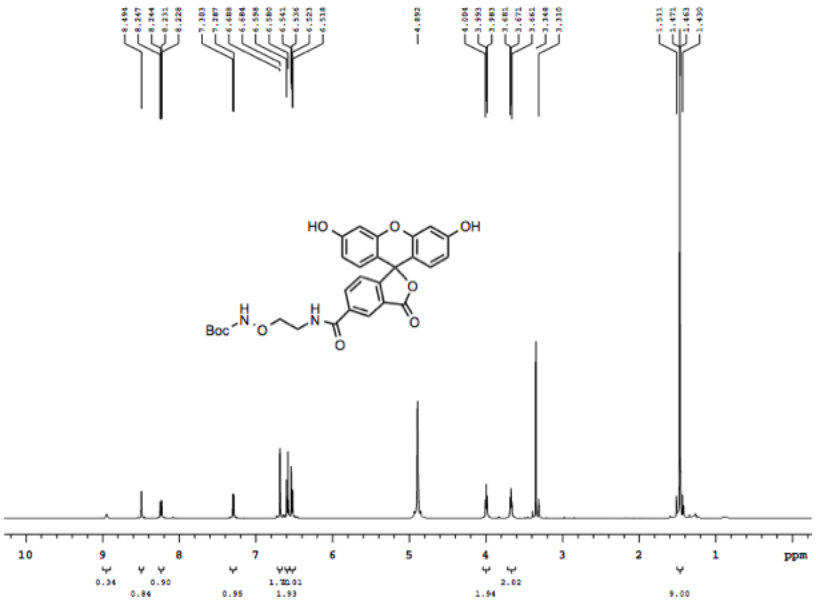
NMR AND MS DATA FOR SECTION 3

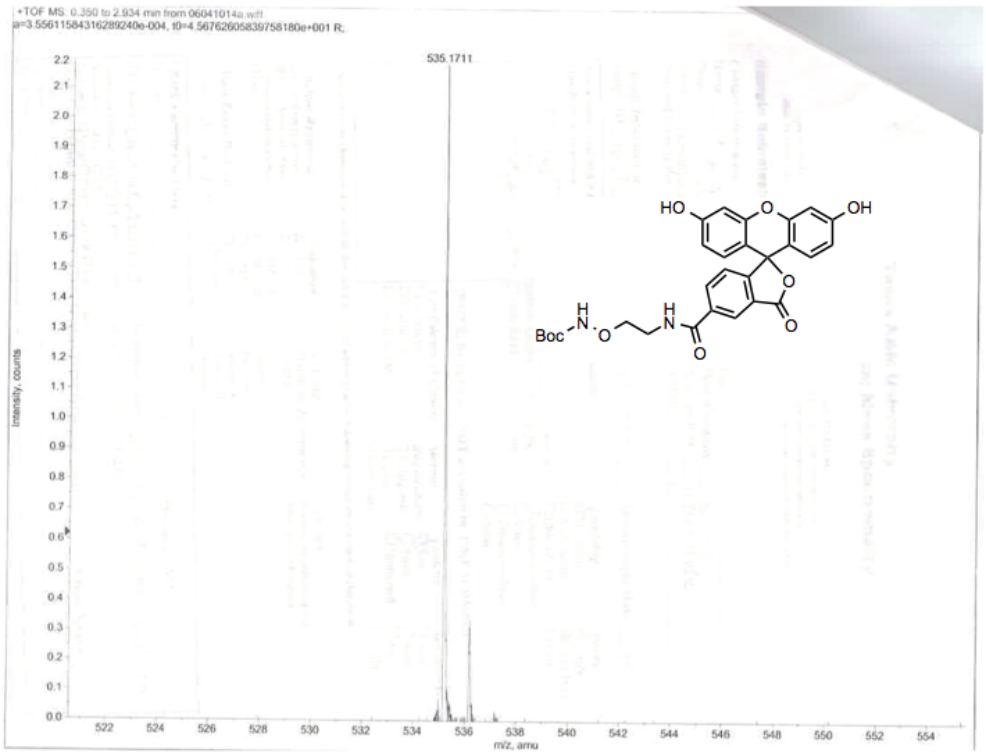


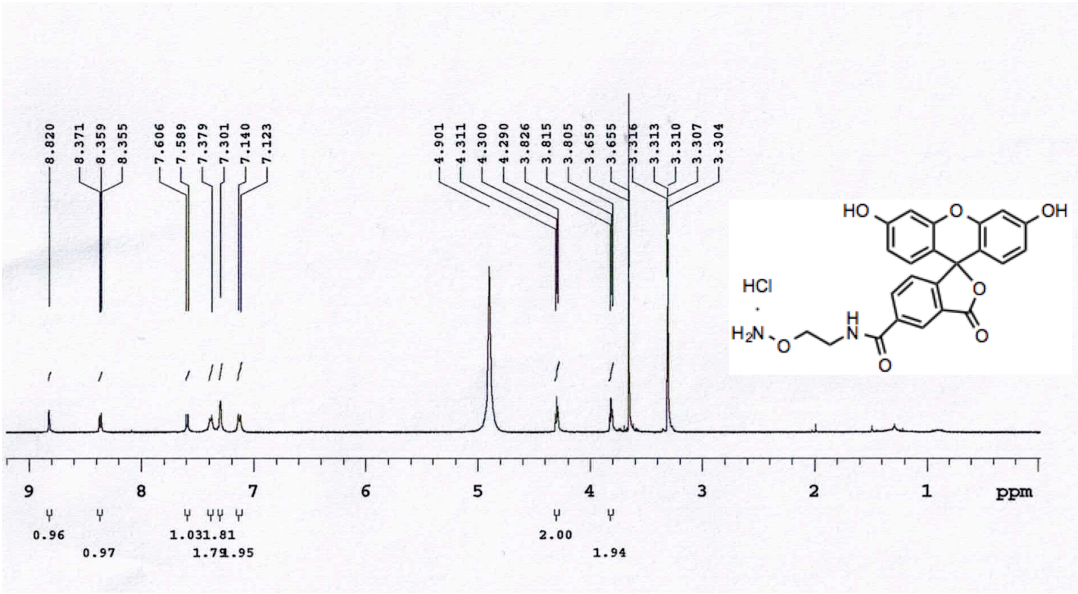

192.082
169.408
167.406
136.686
136.627
136.145
130.468
129.359
129.164
77.160
67.193
63.053
37.594
23.194
14.168
O
H
CO2Et
CO2Et
NH
O
ppm 180
160
140
120
100
80
60
40
20
0

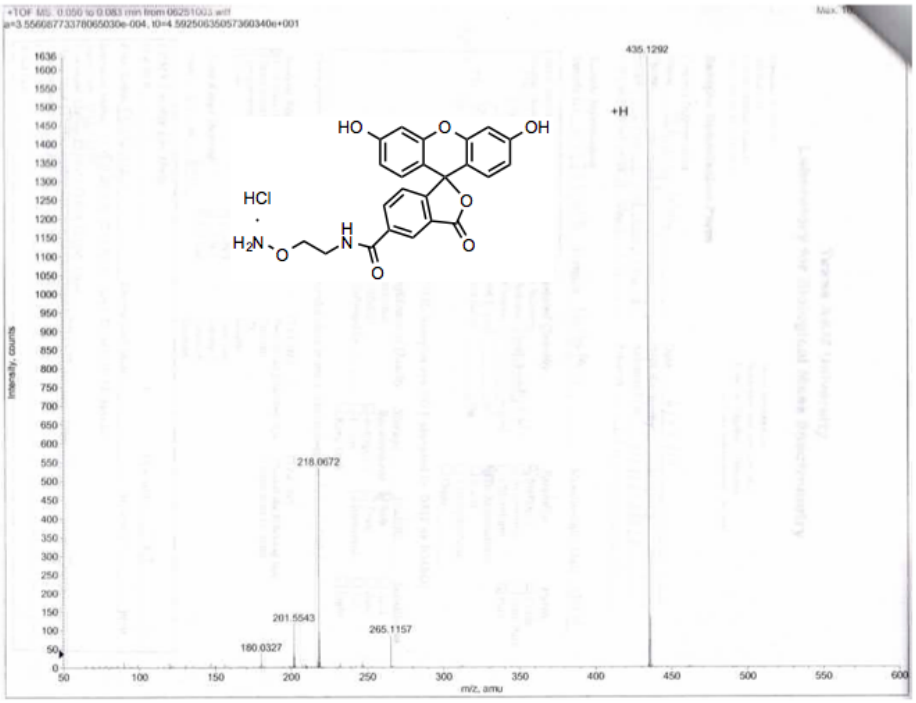

9.799
7.794
7.771
7.721
7.519
4.187
3.337
3.316
3.288
3.269
3.222
3.196
3.173
3.149
O
H
O
OH
NH2
HCl
0.904
1
2.14
0.991
1.13
2.97
ppm 10
9
8
7
6
5
4
3
2
1
0


195.936
171.279
136.165
136.095
135.191
130.322
129.896
129.797
53.922
35.211
O
H
O
OH
NH2
HCl
220
200
180
160
140
120
100
80
60
40
20
0


HO
O
OH
O
O
H
N
Boc
O
H
N
O
10
9
8
7
6
5
4
3
2
1
ppm

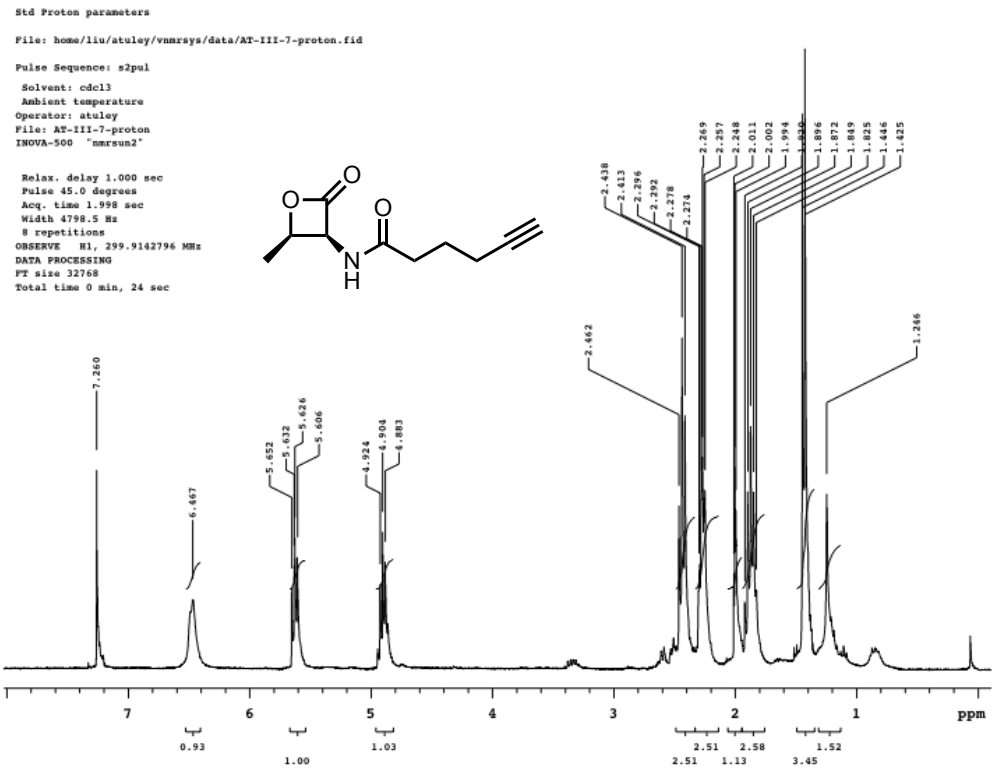

+TOF MS: 0.350 to 2.934 min from 06041014a.wiff
a=3.556115843162892 40e-004, t0=4.567626058397581 80e+001 R;
535.1711
Intensity, counts
m/z, amu
HO
OH
O
O
O
H
N
O
Boc
H
N
O

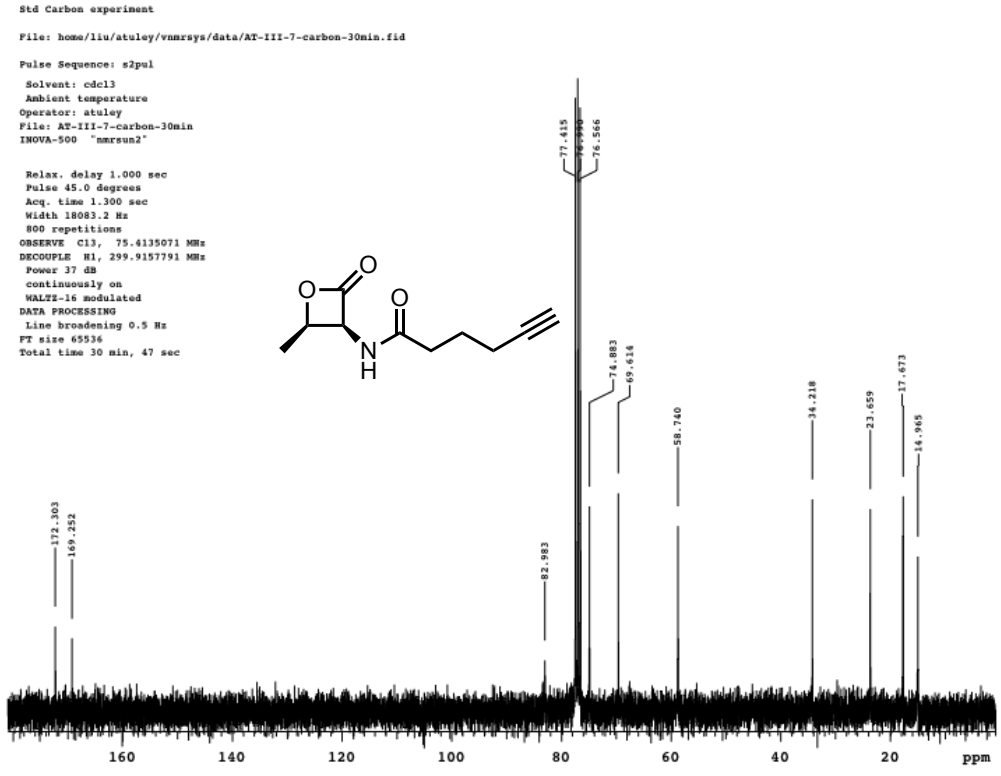

8.820
8.371
8.359
8.355
7.606
7.589
7.379
7.301
7.140
7.123
4.901
4.311
4.300
4.290
3.826
3.815
3.805
3.659
3.655
3.316
3.313
3.310
3.307
3.304
HO
O
OH
HCl
H2N
O
H
N
O
O
O
9
8
7
6
5
4
3
2
1
ppm
0.96
0.97
1.03
1.81
1.79
1.95
2.00
1.94


435.1292
+H
HO
O
OH
HCl
H2N
O
H
N
O
O
O
218.0672
201.5543
265.1157
180.0327
Intensity, counts
m/z, amu

APPENDIX C

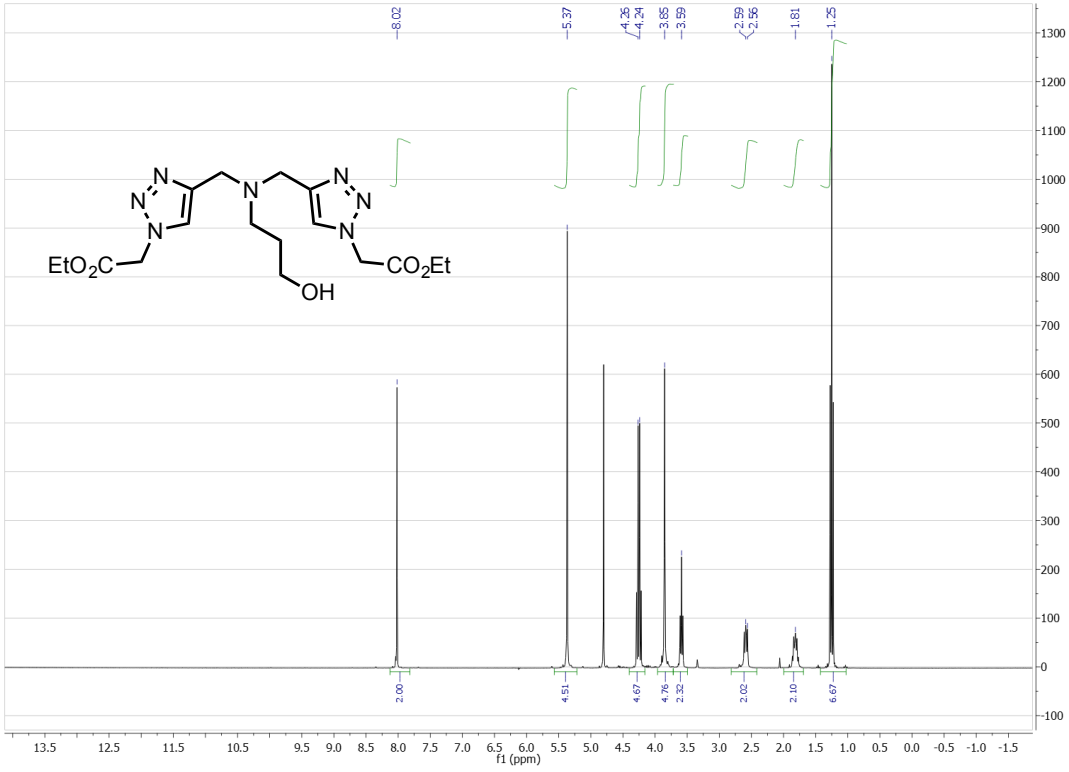
NMR DATA FOR SECTION 4

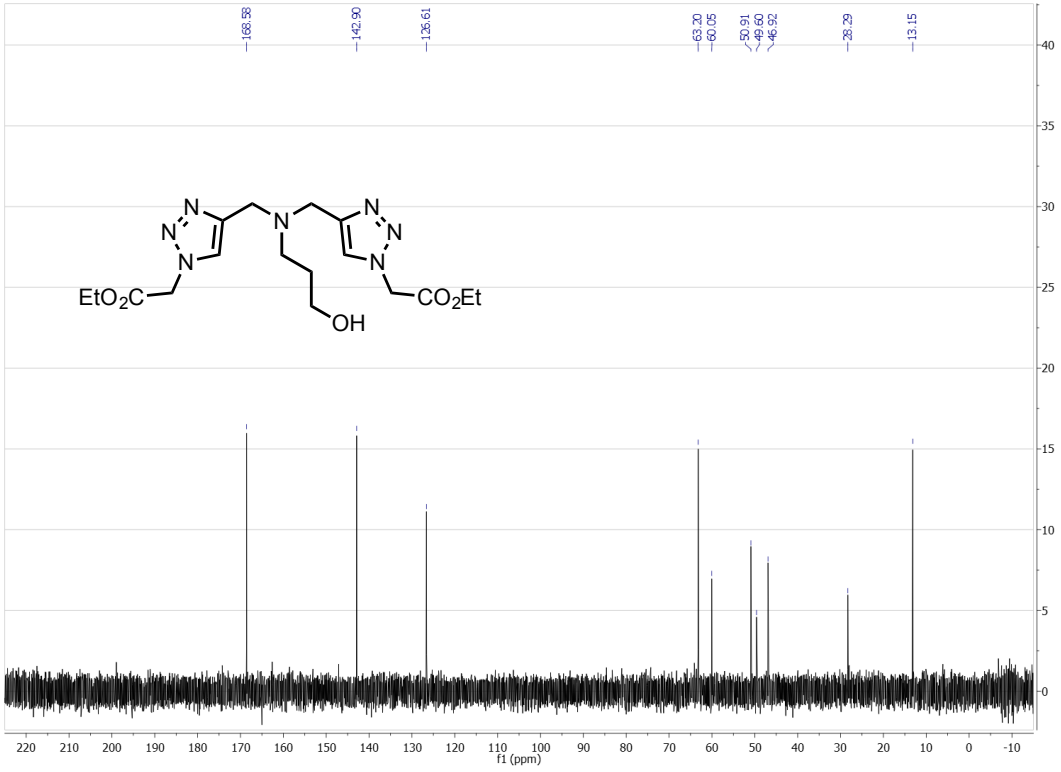


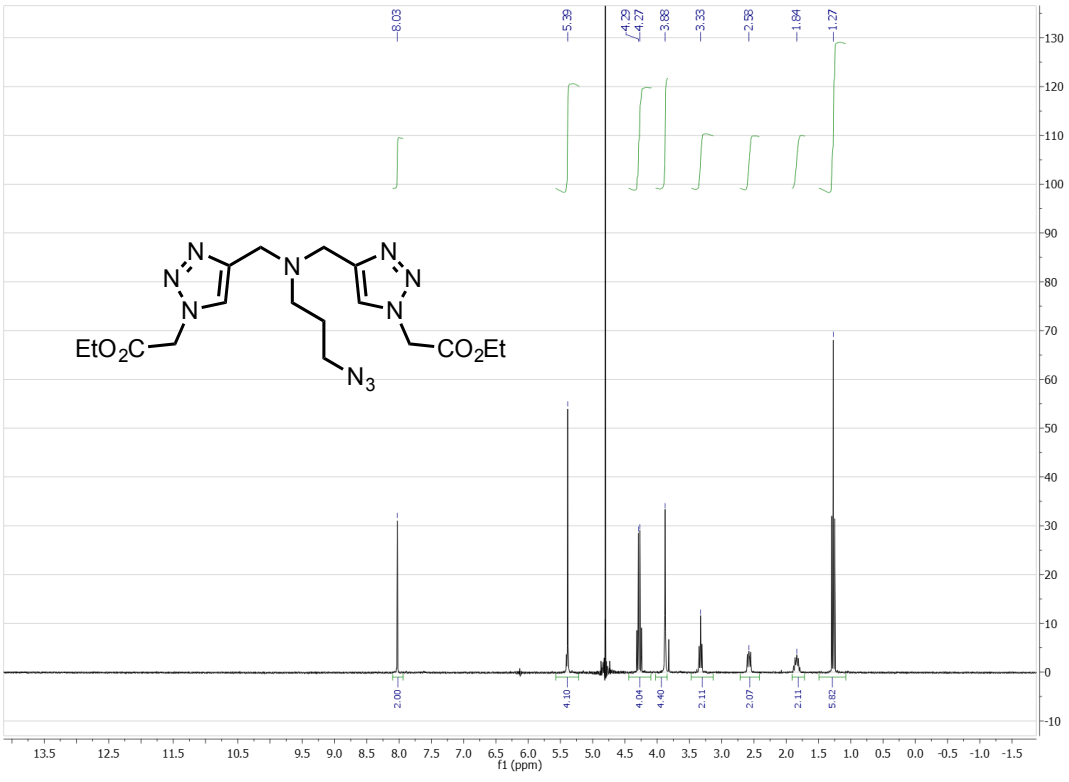

Std Carbon experiment
File: home/liu/atuley/vnmrsys/data/AT-III-7-carbon-30min.fid
Pulse Sequence: s2pul
Solvent: cdcl3
Ambient temperature
Operator: atuley
File: AT-III-7-carbon-30min
INOVA-500 "nmrsun2"
Relax. delay 1.000 sec
Pulse 45.0 degrees
Acq. time 1.300 sec
Width 18083.2 Hz
800 repetitions
OBSERVE C13, 75.4135071 MHz
DECOUPLE H1, 299.9157791 MHz
Power 37 dB
continuously on
WALTZ-16 modulated
DATA PROCESSING
Line broadening 0.5 Hz
FT size 65536
Total time 30 min, 47 sec
O
O
O
N
H
172.303
169.252
82.983
77.415
76.990
76.566
74.883
69.614
58.740
34.218
23.659
17.673
14.965
160
140
120
100
80
60
40
20
ppm

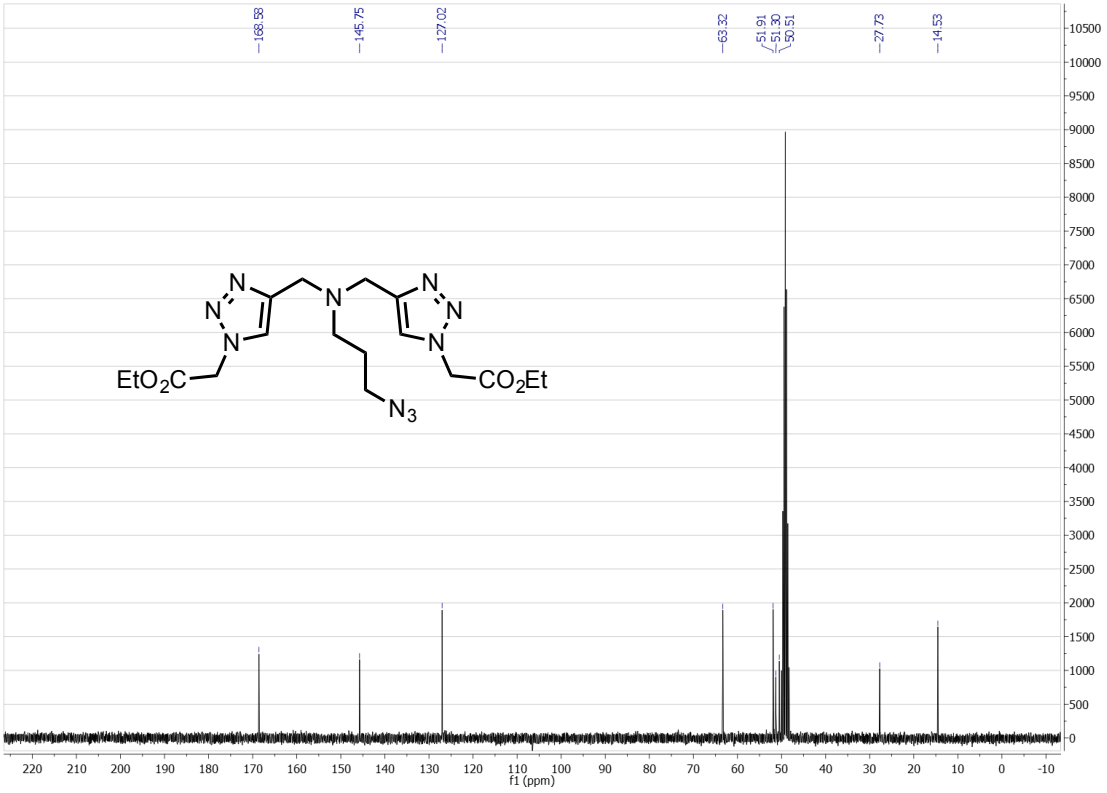
APPENDIX D

NMR DATA FOR SECTION 5




168.38
142.90
136.61
63.20
60.05
50.91
49.60
46.92
28.29
13.15
N
N
N
N
N
N
EtO2C
CO2Et
OH
f1 (ppm)


8.03
5.39
4.29
4.27
3.88
3.33
2.58
1.84
1.27
N
N
N
N
N
N
EtO2C
CO2Et
N3
2.00
4.10
4.04
4.40
2.11
2.07
2.11
5.82
f1 (ppm)


168.88
145.75
127.02
63.32
51.91
51.30
50.51
27.73
14.53
N
N
N
N
N
N
EtO2C
CO2Et
N3
f1 (ppm)